

A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-498 <PAR>
 A:Cross-references: GB:AL162758; GB:AL157959; NID:97380672; PIDN:CB85327.1; PID:9738073
 A:Experimental source: serogroup A, strain 22491
 C:Genetics:
 A:Gene: mafB2; NMA2113

alignment_scores:

Quality:	730.50	Length:	526
Ratio:	2.364	Gaps:	14
Percent Similarity:	38.745	Percent Identity:	34.221

alignment_block:

US-09-303-518d-465 x H81782 ..

Align seg 1/1 to: H81782 from: 1 to: 498

```

28 ATCTGTCATAGTGGAGGTGTC.....CTGCCGAT 59
   ::::::::::::::::::::|||
7  LeuileuLeuAlaAlaCysAlaValAlaAlaAlaLeuileu1 23
   ::::::::::::::::::::|||
60 GCATGCACACGGCTCAGATTGGCAACGATTCTTTATCCGGCAGTTTC 109
   ::::::::::::::::::::|||
23 nProAlaLeuAlaAlaAspLeuAlaGlnAspProPheileThrAspAsnA 40
   ::::::::::::::::::::|||
110 TCGACCGTCAGCATTTGCAACCCGACGGAAATACACCTATTGGCGAGC 159
   ::::::::::::::::::::|||
40 laglaArgGlnHisTyrGluProGlyLysTyrHisLeuPheGlyAsp 56
   ::::::::::::::::::::|||
160 ..AGGGGGGAAGCTTGGCGAGCGCAGCGGTATATCGATTGGGAACAT 206
   ::::::::::::::::::::|||
57 ProAlaGlySerValSerAspArgThrGlyGlnIleAsnValIleGlnAs 73
   ::::::::::::::::::::|||
207 ACAAGCATCATGATGGGCAACCTGTTCATCCAGCGCGCCATTAAAG 256
   ::::::::::::::::::::|||
73 pTyThrHisArgMetGlyAsnLeuileuileGlnIleAsnIleAsnG 90
   ::::::::::::::::::::|||
257 GAATATCGGCTACATGTGCGCTTTTCGATCAGCGGCGCAAGATCCAT 306
   ::::::::::::::::::::|||
90 lYThrIleGlyTyrHisThrArgPheSerGlyHisGlyTyrGlnGlnHis 106
   ::::::::::::::::::::|||
307 TCCCGCTTCGACAAACGATCCCTCAGATTCGATTCGATGACCGCGTAG 356
   ::::::::::::::::::::|||
107 AlaProPheAspAsnHisAlaAlaAspSerAlaSerGlnGlyAsnGlyAs 123
   ::::::::::::::::::::|||
357 TCCCGTTGACGATTCAGCTTTACCGCATTCATGGAGCGATACGAAAC 406
   ::::::::::::::::::::|||
123 nValAspArgLysIlePheThrValTyrArgLeuAsnTrpGlnGlyHisGlu 140
   ::::::::::::::::::::|||
407 ACCATCCCGCGGAGCGCTATAGCGGCGACAGGGCGCGGCTATCCGCT 456
   ::::::::::::::::::::|||
140 lshAspProAlaAspAlaTyrAspGlyProLysGlyAsnTyrProLys 156
   ::::::::::::::::::::|||
457 CCCAAAGGCGGAGGATATATACGATACGATACGATTAAGCGCTGCCA 506
   ::::::::::::::::::::|||
157 ProThrGlyAlaAlaArgAspGlyTyrThrTyrHisValAsnGlyThrAla 173
   ::::::::::::::::::::|||
507 AAATATCGCGCTCAACCTGACGCAACACGACGCGCAACGCGCTG 556
   ::::::::::::::::::::|||
173 gSerIleLysLeuAsnProThrAspThrArgSerIleArgGlnArgLys 190
   ::::::::::::::::::::|||
557 TCGACCGTTTCACATACGCGTAGTATGCTGACGCAAGAGAGGCGAC 606
   ::::::::::::::::::::|||
190 eArAspAsnTyrSerAsnLeuGlySerAsnPheSerAspArgAlaAspGlu 206
   ::::::::::::::::::::|||
607 GGATTCAAGCGCGCACCGATACAGCCGAGTGACGAGATCGGGCAA 656
   ::::::::::::::::::::|||
207 AlaAsnArgLysMetPheGlnHisAsnAlaLysLeuAspArgTyrTrpLys 223
   ::::::::::::::::::::|||
657 TGGCGCGGAAGCTTTCAAGCGCAGTCAATATCGTCAAAAACATCATCG 706
   ::::::::::::::::::::|||

```

```

223 nSerMetGluPheIleAsnGlyValAlaAlaGlyAlaLeuAsnProPheI 240
   ::::::::::::::::::::|||
707 GCGCGGAGAGAAATGTCGCGCCAGCGCATCGCGTGAAGGTATAGC 756
   ::::::::::::::::::::|||
240 lSerAlaGlyGlyAlaLeuGlyIleGlyAspIleLeuTyrGlyThrArg 256
   ::::::::::::::::::::|||
757 GAAGGTCGAAACATTTGCTTTATGACAGCGCTGGGCTGCTTCACCGG 806
   ::::::::::::::::::::|||
257 TyrAlaIleAspLysAlaAlaMetArgAsnIleAlaProLeuProAlaG 273
   ::::::::::::::::::::|||
807 AACACAGATGCGCGCATCAGCATTTGGCAGATATGCGCAACGCAAG 856
   ::::::::::::::::::::|||
273 uGlyLysPheAlaValIleGlyLeuGlySerValAlaGlyPheGln 290
   ::::::::::::::::::::|||
857 ACTATGCGCGAGACGCTCCGATTCGGCAGTCCAAAACCAATGCC 906
   ::::::::::::::::::::|||
290 yAsnThrArgGluAlaValAspArgTrpIleGlnGlnAsnProAsnAla 306
   ::::::::::::::::::::|||
907 GCACAAAGCATAGAGCGCTGACGATATCTTACGCGAGTTCGCCGT 956
   ::::::::::::::::::::|||
307 AlaGlnThrValGluAlaLeuValAsn.....ValLeuProPhe 319
   ::::::::::::::::::::|||
957 CAAGGATTTGAGCTGTTGCGGAAATACGCTTGGCGGCATCAGCG 1006
   ::::::::::::::::::::|||
319 e..... 319
   ::::::::::::::::::::|||
1007 CACATCCTGTCACACCGCTGCCAGATGGCGAGATCCGATGCCGAAGG 1056
   ::::::::::::::::::::|||
320 .....AlaLysValLysAsnLeuThrLysAlaAlaLysPro...Gly 332
   ::::::::::::::::::::|||
1057 AAATCGCGCTCAGCGACAAATTTGCCGATGGCGCATACGCCAAATATCC 1106
   ::::::::::::::::::::|||
333 LysAlaAlaValSerGlyAspPheSerAlaAla..... 343
   ::::::::::::::::::::|||
1107 GTCCCTTACATTCGCCGAATATCCGT..... 1134
   ::::::::::::::::::::|||
344 .....TyrAsnThrArgThrThrArgLysValThrThrGluThrGluG 358
   ::::::::::::::::::::|||
1135 .....TCAACTTG...GAGCAG 1149
   ::::::::::::::::::::|||
358 lYeuAsnArgIleArgGlnAsnGlnLysAsnSerAsnIleHisGluLys 374
   ::::::::::::::::::::|||
1150 CGTTACGCAAGAAACATCCTCCTCAGACCGCGCGCTCAAC... 1197
   ::::::::::::::::::::|||
375 AsnTyrGlyArgAspAsnProAsnHisIleAsnValLeuSerGlyAsnSe 391
   ::::::::::::::::::::|||
1197 ..... 1197
   ::::::::::::::::::::|||
391 rIleGlnHisIleLeuTyrGlyAspGluAlaGlyGlyHisLeuPheP 408
   ::::::::::::::::::::|||
1198 ..GAAAGAATGTGAACCTGGCAACAAACGCCAC..... 1230
   ::::::::::::::::::::|||
408 roGlyLysProGlyLysThrThrPheProGlnHisTrpSerAlaSerLys 424
   ::::::::::::::::::::|||
1231 .....CGAAGACCAAA..... 1242
   ::::::::::::::::::::|||
425 lIeThrHisGluIleSerAspIleValThrSerProLysThrGlnTrpLys 441
   ::::::::::::::::::::|||
1243 .....GTCCGTTGACGCGTGAAGGTTTCGCAATT 1273
   ::::::::::::::::::::|||
441 rAlaGlnThrGlyThrGlyLysTyrIleAlaLysGlyArgPro.... 456
   ::::::::::::::::::::|||
1274 TTGAAAAGACGTAATAATGATACGAGAG..... 1302
   ::::::::::::::::::::|||
457 ..AlaArgTrpValSerTyrGluThrArgAspGlyIleArgIleArgThr 472
   ::::::::::::::::::::|||
1303 .....ATTAATACCGGTGTACCAAGTGA 1328
   ::::::::::::::::::::|||
473 ValTyrGluProAlaThrGlyLysValAlaThrAlaPhePro.....As 487
   ::::::::::::::::::::|||
1329 TCCTATGATGAACCGCTTTAATCT 1356
   ::::::::::::::::::::|||
487 pArgThrSerAsnProLysTyrAsnPro 496

```

```
seq_name: p1r2:B82028
seq_documentation_block:
adhesin NMA0324 [imported] - Neisseria meningitidis (strain Z2491 serogroup A)
C:Species: Neisseria meningitidis
C:Date: 05-May-2000 #sequence_revision 05-May-2000 #text_change 02-Feb-2001
C:Accession: B82028
R:Parhill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; Morel,
R.; Parhill, S.; Jørgensen, K.; Leather, S.; Moule, S.; Mungall, K.; Quail, M.A.; Rajadream,
Nature 404, 502-506, 2000
A:Title: Complete DNA sequence of a serogroup A strain of Neisseria meningitidis Z2491.
A:Reference number: A81775; MUID:20222556
A:Status: preliminary
A:Accession: B82028
A:Molecule type: DNA
A:Residues: 1482 <PAR>
A:Cross-references: GB:AL162752; GB:AL157959; NID:g7378778; PIDN:CA883629.1; PID:g737908
A:Experimental source: serogroup A, strain Z2491
C:Genetics:
A:Gene: mafB; NMA0324
```

```
alignment_scores:
Quality: 419.50 Length: 447
Ratio: 1.665 Gaps: 18
Percent Similarity: 56.376 Percent Identity: 31.320
```

```
alignment_block:
US-09-303-518D-465 x B82028 ..
```

```
Align seg 1/1 to: B82028 from: 1 to: 482
```

```
31 CTGTCATCTAGCGATGTCGTCGCGATGATGACACACCGCTCAGATT 80
:::|||||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
16 lIleSerLeuLeuGlnIleProIle.....SerHisAlaasnGlyLe 29
44 roGlyGlyLysTyrHisLeuPheGlyAsnAlaArgLysValLysAsn 60
81 GGCAACGATCTTTTATCCGGCAGGTTTCGACCGTCACGATTTGCAAC 130
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
29 u.....AspAlaArgLeuArgAspAspMetGlnAlaLysHisTyrGlu 44
131 CCGACGGGAATACCACTATTCGGC...AGCAGGGGGGAACCTTGGCGAG 177
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
44 roGlyGlyLysTyrHisLeuPheGlyAsnAlaArgLysValLysAsn 60
178 CGCAGCGCATATCGGATGGGAAACATACAAAGCCATCATGTTGGGCAA 227
|||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
61 ArgVal.....TyrAlaValGlnThrPheAspAlaThrAlaValGlyPr 75
228 CTTGTTTC.....ATCCAGCAGGGGCCATTAAAGAAATATCGCT 268
:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
75 oIleLeuProIleThrHisGlnArgThrGlyPheGlnGlyLeIleGlyT 92
269 ACATGTCGCGCTTTCGATCAGGGGACAGAGTCATTCGCCCTTGCAG 318
|||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
92 yrgLutThrHisPheSerGlyHisGlyHisGlyValHisSerProPheAsp 108
319 AACCATCCGCACATTCGATTCGATGAGACCGGTCGATCCGCTGACGG 368
|||||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
109 AsnHisAspSerLysSerThrSerAspPheSerGlyGlyValAspGlyG 125
369 ATTCCAGCTTTACCGCATTCATCGGACGAGATAGACACCATCCGCGCG 418
|||||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
125 yPheThrValTyrGlnLeuHisArgThrGlySerGlnIleHisProGlnA 142
419 ACGGCTATGACGGGACAGGGGCGCGGATATCCCGCTCCCAAGCGCG 468
|||||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
142 spGlyTyrAspGlyProGlnGlySerAspTyrProProGlyGlyAla 158
469 AGGATATATACAGCTACGACATTAAGGGGTCGCCCAAAATATCCGCT 518
|||||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
159 ArgAspIleTyrSerTyrTyrValLysGlyThrSerThrLysThrLysSe 175
```

```
519 CAACCTGACCGCAACCGCAGCCGACCAACAGCGCTTGCAGCTTCC 568
|||||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
175 rAsnIleValIProArgAlaProPheSerAspArgThrPheLysGlnAsnA 192
569 ACAATACCGGTGATTCCTGACCGCAAGATAGCGGACGATTCAAAGC 618
:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
192 lAgIyAlaIleSerGlyPhePheSerArgAlaAspGlnAlaGlyLysLeu 208
619 GCCACCCGATACGCCGCCGATGAGATGAGATGCGGCATGCGCGGAGC 668
:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
209 lIeThrPheSerAspProAsnLysAsnThrPheAlaValSerProValThr 225
669 TTTCACGCGCACTGCAATATGTCAAAAC.....ATCATCGCGCGG 712
:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
225 pIleArgGlyIleValGlnGlyAlaValAsnProPheLeuMetGlyPheG 242
713 CAGGAGAAATTCGCGCGCA.....GGCGATCCGCTGACGGGTAAAC 756
|||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
242 lGnIyValGlyIleGlyAlaIleThrAspSerAlaValSerProValThr 258
757 GAAGGCTCAACATTCGTCGTATGACGCG.....TTGGTCTGCT 797
:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
259 AspThrAlaIleGlnGlnThrLeuGlnGlyIleAsnHisLeuGlyAsnLe 275
798 TTCCACCGAAACAAAGATGCGCGCATCAGCATTTGGCAGATATGCGG 846
|||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
275 uSerProGlnAlaGlnLeuAlaIleAlaThrAlaLeuGlnAspSerAla 292
847 ..CAACTCAAGACTATGCGCGCAGCAGCATTCGGATTTGGCACTCCA 894
:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
292 heAlaValLysAspGlyIleAsnSerAla...ArgGlnThrAlaAspAla 307
895 AACCCCAAT.....GCCGCAACAGCATAGAAAGCCGTCAGCAATAT 935
:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
308 HisProAsnIleThrAlaThrAlaGlnThrAlaLeuAlaValAlaGlnAl 324
936 CTTTACGCGCATTCATCCC.....GCTTTCGGGGGAAA 954
|||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
324 aaIaThrThrValITrpgLylLysLysValGlnLeuAsnProThrLysT 341
955 .....GTCAAGGATTCGA.....GCTTTCGGGGGAAA 984
|||||:::|:::|:::|:::|:::|:::|:::|:::|:::|
341 rPAspTrpAlaLysAsnThrGlyTyrLysThrProAlaValArg..... 355
985 TAGCGCTTGGCGGCATCAGGACATCCTGTCAGACGCGTCAGATGG 1034
|||:::|:::|:::|:::|:::|:::|:::|:::|:::|
356 .....ThmethIsthrLeuAspGlyGlnMetAlaG 366
1035 CGAGATCGCATTCGCGAAAGGAAATCCGCGTCAGCAGCAATTTGCGG 1084
|:::|:::|:::|:::|:::|:::|:::|:::|:::|
366 yGlyAsnArgProPro.....LysSerIleThrSerAsnSerLysAlaA 381
1085 ATGCG..... 1089
|||||
381 sPAlaSerThrGlnProSerLeuGlnAlaGlnLeuIleGlyGlnGlnIle 397
1090 .....GCATACGCCAATATC.....CCGTCCCTTACCA 1118
|||||:::|:::|:::|:::|:::|:::|:::|:::|
398 SerSerGlyHisAlaTyrAsnLysHisValIleArgGlnGlnGlnPheTh 414
1119 TTCCGGAATATTCGTTCAAACTTGAGCAGCAGCGTTACGGCAAAACA 1168
:::|:::|:::|:::|:::|:::|:::|:::|:::|
414 rAspLeuAsnIleAsnSerProAlaAspPheAlaArgHisIleGlnAsnI 431
1169 TCACCTCTCAACCGTCCCGCGTCAAACGAAAGAAATGATGT 1209
|||:::|:::|:::|:::|:::|:::|:::|:::|
431 lValSerHis.....ProThrAsnMetLysGlnLeu 441
```

```
seq_name: p1r2:T4814
```

```
seq_documentation_block:
hypotheical protein 1566.220 [imported] - Neurospora crassa
C:Species: Neurospora crassa
```

C>Date: 05-May-2000 #sequence_revision 05-May-2000 #text_change 05-May-2000
C/Accession: T48814
R/Schulte, U.; Aign, V.; Hobeisel, J.; Brandt, P.; Fartmann, B.; Holland, R.; Nyakatura,
submitted to the Protein Sequence Database, April 2000
A/Reference number: 224541
A/Accession: T48814
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-1952 <SCH>
A/Cross-references: EMBL:AL353822; GSPDB:GN00112; NCSP:1556.220
A/Experimental source: cosmid contig 15E6; strain 74
C/Genetics:
A/Gene: NCSP:1556.220
A/Map position: 2
A/Introns: 281/3

alignment_scores:
Quality: 170.00 Length: 514
Ratio: 0.742 Gaps: 24
Percent similarity: 44.553 Percent identity: 23.735

alignment_block:
US-09-303-518d-465 x T48814 ..
Align seg 1/1 to: T48814 from: 1 to: 1952

```
125 TCGAACCCGACGGAGAAATACCACTTATCGCAGCAGGGGGAACCTTGGC 174
      |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
1221 AATThPrOlaAlaProValHsAlaThPrOlaThPrOlaThPrOlaThPrOla 1237
175 GAGCCGACGGGTCATATCGGATTGGGAACATCAAAAGCCATCAGTTGGG 224
      |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
1237 OAla.....AsnHsHsThrMetAlaProA 1246
225 CAACCTGTCATCCAGCAGCGGCAATTAAGAAATATCGGTACATG 274
      |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
1246 lAThAlaProSerAlaAlaGlnAlaThPrOlaThPrOlaThPrOlaThPrOla 1262
275 .....TCCGCTTTCGATCCAGCGGCA 297
1263 ArgAlaAsnProProGlyGlyAsnProSerAlaValAlaProArgThrIleTh 1279
298 GAAGTCATCCCTTCGACAAACATGCTCAGATT.....C 335
1279 rAlaAlaAlaProAlaProThPrOlaProLeuProAlaProAlaAla 1296
336 CGATTCTGATGAAGCGGTAGTCCGTTGACGGATTCAGCCTTACCGCA 385
      |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
1296 rOleSerIyHsProIleAlaProSerThrValAsnAsnThrAla 1312
386 TCCATTGGGAGATACGAACACCATCCGCGAGCGGTATGACGGGC.. 433
      |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
1313 AlameGlyAlaArgPro.....ValProAlaAlaValAlaProGlyPr 1326
434 .....CACAGGGCGGGCTATCCCGCTC..... 457
1326 oGlnIleGlyAlaHsSerGlyAlaAlaIleGlyAlaThPrAlaProValS 1343
458 .....CCAAAGCGCGAGGATATATACAGCTACGACATTAAGA 496
      |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
1343 eArgSerProProValThrHsIleGlyAlaValAlaThPrAlaAla..... 1356
497 GCGTTGCCCAAAATATCCGCTCAACCTGACGACAAACGCGACACCGGA 546
      |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
1357 .....ProThrThrThrProArgse 1363
547 CAACGCTTGTGACCGTT.....T 566
      |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
1363 rAsnIleuAlaAlaSerValProAlaGlySerMetAlaGlnIleThrHs 1380
567 CCACATATACCGGTAGTATGCTGACGCAAGAGTAGCGAGGATTCGAAC 616
      |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
```

```
1380 eArgAlaProAla.....GlySerGlyAsnAlaSerIleSerArg 1393
617 GCGCCACCGCATACAGCCCGACAGTGCAGATCGGCAATGCCCGCA 666
      |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
1394 AlaProSerThrAlaAlaProThPrOlaProThSerValProProVal 1410
667 GCTTTCACGGCAGCTGCAGATATCGTCAAAACATCATCGCGCGCAG 716
      |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
1410 lSerSerThrValArgProMetSerIyValProThrGly.....Prop 1425
717 AGAAATTGTCCGCGCAGGGGATGCCGTGCAGAGGTATACGAAAGCTCA 766
      |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
1425 rOgIyLeuThrAla..... 1429
767 ACATTGCTGTATGACAGGCTTGGTCTGCTTTCACCGCAAAACAGATG 816
      |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
1430 .....ProProThrAlaSerSe 1435
817 GCGCGCATACAGATTGGCAGATATGGCGCAACTCAAGACTATGCCG 865
      |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
1435 rGlyAlaSerAlaGlyTyrAlaArgProAlaAsnAlaSerThrMetProA 1452
866 .....CAGCAGCCATCCGCGATTGGCGATGCCGCAAA 895
1452 lAProThrSerGlyAlaAlaAlaMetThSerAlaValProGlnSer... 1467
896 ACCCCAAATGCCGACAAAGCATAGAACCCGTCAGCAATATCTTTACGCA 945
      |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
1468 ValProSerPro.....ArgProSerSerLeuThrIleGly 1480
946 GTCATCCCGCTCAAGAGGATTGGAGCTGTTCCGGGCAAAATACGGCTGG 995
      |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
1480 n.....MetGlyValProAlaIleAla 1488
996 CGGATCAGCGCACATCTGTCAAGCGGTGCGCAGATGGCGAGATGCAT 1045
      |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
1488 lAlaSerThrSerArgProAlaSerGly..... 1497
1046 TCCCGAAAGGAATCCGCGTCAGCAGCAATTTGCGATGCGGAT.. 1093
      |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
1498 .....ValTyrAsnProProAlaSerSerLeuAlaProSerThrHsIy 1512
1094 ....ACGCCAANTACCCGCTTCCATTCGCCAATATCCGTTCAAA 1139
      |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
1512 sSerMetProSerAlaValAlaProThrThrAlaSerGlyAlaValSerSeT 1529
1140 CTGAGCAGCGCTTACGCGAAAGAAACATCACTTCACCGTGS.... 1185
      |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
1529 hValSerSerLeuAlaAlaThrProLeuProProSerAlaProArgTyr 1545
1186 CCGCGCTCAACGGAAGATGTGAAACTGGCAMAACCAAGCCACCGCA 1235
      |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
1546 GlyProSerAsnAsnAlaThrProThrThrProAlaThAlaIleProAr 1562
1236 GACCAAAAGTCCGTTGACGTA.....GGGTTTC 1267
      |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
1562 gAlaProLeuProAlaAlaSerSerValSerAlaProValThrGlyGlnP 1579
1268 CGAATTTTGAAGAGAGTAAATACGATACGAGATCAAGATTAACCGCTGA 1317
      |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
1579 rOserPhSerAlaProAlaSerValProThrProProThrSerAlaAla 1595
1318 CCACAA...GTGAATCTATTAAGTGAACCGGTCTTAATCT... 1356
      |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
1596 SerGlnGlyAlaGlnProLeuSerGlnProVal...GlyProAlaProAl 1611
1357 ....AAAGTCTGTGCGATCGGCTCATCTTGCTATTAACGCGCAGAA 1402
      |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
1611 AlIeSerSerSerIleGlyAlaThrProAlaAlaSerIleProSerSeA 1628
1403 TTCAATATCCCAAAATTTACCAAGCAAGTAGAATCATAT...ATCCCA 1449
      |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
1628 lAProAlaAlaIleuAlaPro.....ValThrTyrProValPro 1640
```

1450 CCTAAATTAATCTCTCTCAGACCGCTACCAAGGACCT 1491
 1641 GlnGlnAlaSerAlaAlaAlaAlaArgLeuProValThrPro 1654
 seq_name: p1r2:T45025

seq_documentation_block:
 mucin MUC5B, tracheobronchial [imported] - human (fragment)
 C:Species: Homo sapiens (man)
 C:Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 21-Jul-2000
 C:Accession: T45025
 R:Desseyn, J.L.; Guyonnet-Duperrat, V.; Porchet, N.; Aubert, J.P.; Laine, A.
 A:Title: Human mucin gene MUC5B, the 10.7 kb large central exon encodes various alternat
 A:Reference number: 222899; MUID:97166151
 A:Accession: T45025
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-3570 <DESeq>
 A:Cross-references: EMBL:Z72496; NID:g1834502; PIDN:CAA96577.1; PID:g1834503
 A:Experimental source: placenta
 C:Genetics:
 A:Gene: MUC5B

alignment_scores:
 Quality: 159.50 Length: 430
 Ratio: 0.794 Gaps: 20
 Percent Similarity: 46.744 Percent Identity: 23.721

alignment_block:
 US-09-303-518D-465 x T45025 ..

Align seg 1/1 to: T45025 from: 1 to: 3570

12 CGCAAAATATCCCTATCTCTCCATCTGCGAGGTCCTCCGCGATGC 61
 1197 ProSerSerThrProGlyThrAlaHisThrLysValProThrThrTh 3213
 62 ATGCAC.....A 68
 3213 ThrThrGlyPheThrAlaThrProSerSerProGlyThrAlaLeu 3230
 69 CGCCTCAGATTGGCAAGCATCTTTATCGCGAGGTTCTCGACGCTC 118
 3230 hrProProValThrPheSerThrThrThrProThrThrThrPro 3246
 119 AGCATTTGAAACCGCGGAAATACACCTATTGCGAGCGAGGCGAA 168
 3247 ThrThrSerGlySerThr.....ValThrProSerSerIleProGlyTh 3261
 169 CTGGCGGAGCGACGCTCATATGCGATTGGAAACATACAAAGCCATC 217
 3261 rThHisThrAlaArgValLeuThrThrThrThrValAlaThrG 3278
 218 ..AGTTGGCAACCTGTTCAATCAGACGCGCGCATTA.....AAG 256
 3278 LysSerMetAlaThrProSerSerThrThrThrSerGlyThrPro 3294
 257 GAAATATCGGCTACATGTCGCTTTCCATCAGCGGACGAGATCCAT 306
 3295 SerLeuThrThrThrAlaThrThrThrAlaThrGlySerThrThr 3311
 307 TCCCGCTTCGACA.....ACATGCTCAACATCCGATCTG.....A 344
 3311 nProSerSerThrProGlyThrThrProIleProProValLeuThrSer 3328
 345 TGAAGCCGAGTAGTCCGTTGACGAGATCAACCTTACCGCATCATGG 394
 3328 eTalThrThrProAlaAlaThrSerSerLysAlaThrSerSerSer 3344
 395 ACGGATACG...AACACATCCGCGGACGAGGCTATAGCGGCCACAGGGC 441

3345 ProArgThrAlaThrThrLeuProValLeuThrSerThrAlaThrLys 3361
 442 GGGCGCT.....ATCCGCTCCCAAGCGCGAGGATATATACAG 482
 3361 rThrAlaThrSerPheThrThrProIleProSerSerThr...LeuThr 3377
 483 CTACGACATAAAGGCG.....TTGCCAAATATCCGCTCCACAC 523
 3377 hrThrPheValProAlaGlnThrThrThrProMetSerThrMetSer 3393
 524 TGACGACACCGCGACGACCGGAC...AAGGCTTGTGACGCTTTCAC 570
 3394 IleHisThrSerSerThrProGlyThrThrThrThrSerThrVal 3410
 571 AATACCGTAGTATGCTGACGACGAGAGGAGCGGATTCAAAGCGC 620
 3410 rThrThrAlaThrMetThrArg.....AlaThrThrSerThrAla 3424
 621 CACCGGATACAGCCCGAGCTGACAGATCGGCAATGCCCGAAGCTT 670
 3424 hrProSerSerThrLeuGlyThrThr.....ArgLe 3434
 671 TCAACGCGACGACAGATATCGCAAAACATCATCGCGCGCGAGAGAA 720
 3435 LeuThrGluLeuThrThrThrAlaThrThrThrAlaAlaThrGlySer 3451
 721 ATGTGCGCGACGAGGATCCGTCAGGGTATTAAGCAAGGCTCAACAT 770
 3451 rAlaThrLeuSerSerThrPro..... 3458
 771 TGCCTTATGACGCGCTTGCGTCTCTTCCACCGAAACAGATGCGCG 820
 3459GlyThrThrThrPheLeuThrGluPro..... 3467
 821 GCATCAAGATTTGGCAGATATGCGCAGACACCAAGCATATGCCGCGA 870
 3468 ...SerThrIleAlaThrValMetValProThrGlySerThrAlaThr 3483
 871 GCCATCCGCGATTTGGCGATCCCAAAACCCCAATGCCGACAGGCAATGA 920
 3483 rSerSerThrLeuGlyThrAlaHisThr..... 3492
 921 AGCCGTACGAAATATCTTACGGGAGCATCCCGTCAAGGATTTGAG 970
 3493 ..ProLysValAlaThrAlaMetAlaThrMetPro..... 3503
 971 CTGTTCGGGAAATATACGCTTGGCGGATCAAGGACATCCTGTCAAG 1020
 3504ThrAlaThrAlaSerThr..... 3509
 1021 CGGTCCGAGATGGGCGAGATCGCATTCGGAAGGAAATCCGCGTCAG 1070
 3510ValProSer 3513
 1071 CGACAAATTTGGCGATGCGGATACGCAATATCCGCTTACCATT 1120
 3513 eSerThrValGlyThrThrAlaGlnThrProAlaValLeuProSerSer 3529
 1121 CCGCAATATCGCTTCAAACTGGAGCAGC.....GTTAGCGGCAA 1161
 3530 ProThrPheSerValSerThrValSerSerValLeuThrThrLeu 3546
 1162 GAAACATCACTCCCTCA.....CCGTCG 1186
 3546 gProThrGlyPheProSerSerHisPheSerThrProCys 3559

seq_name: p1r2:T30826
 seq_documentation_block:
 nascent polypeptide-associated complex alpha chain, muscle splice form gp220 - mouse
 N:Alternate names: alpha-MNC protein
 C:Species: Mus musculus (house mouse)

C>Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 24-Nov-1999
C/Accession: T30826
R/Yotov, W.V.: St-Arnaud, R
Genes Dev. 10, 1763-1772, 1996
A>Title: Differential splicing in of a proline-rich exon converts alphaNAC into a muscle
A/Reference number: 220889; MUID:96312450
A/Accession: T30826
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: DNA
A/Residues: 1-2187 <YOT>
A/Cross-references: EMBL:U48363; NID:q166668; PID:q1666689; PIDN:AMB18732.1
C/Genetics:
A:Gene: Naca
A:Map position: 10
A:Introns: 24/1; 1996/1; 2024/3; 2050/3; 2099/3; 2142/3; 2183/3
A/Note: differential splicing converts alphaNAC into a tissue-specific DNA-binding actly
C/Keywords: alternative splicing; DNA binding; transcription factor

alignment_scores:

Quality: 156.50 Length: 567
Ratio: 0.660 Gaps: 24
Percent Similarity: 41.799 Percent Identity: 20.988

alignment_block:

US-09-303-518D-465 x T30826 ..

Align seg 1/1 to: T30826 from: 1 to: 2187

```
167 AACCTGGCGAGCGGCGTCATCGATTGGGAACATTAACACCAT 216
    :::::::::::::::::::: ||| ::||| ::||| ::|||
890 SerIleProLysValThrSerProSerProGlnThrProLysSerVa 906
    ::||| ::||| ::||| ::||| ::||| ::||| ::|||
217 CAGT.....TGGCAACCTGTCATCCAGCAGCGCCATTAAGGAA 260
    ::||| ::||| ::||| ::||| ::||| ::||| ::|||
906 IserLeuLysGlyAlaProAlaMetThrSerLys...AlaThrGlu 922
    ::||| ::||| ::||| ::||| ::||| ::||| ::|||
261 TATGGCTACATGTCGCTTCCTTCGATCAGCGCAGCAAGTCCATCCC 310
    ::||| ::||| ::||| ::||| ::||| ::||| ::|||
922 LeuAlaLysSerLysAspValSerProSerGlnPheProLysGluValPro 938
    ::||| ::||| ::||| ::||| ::||| ::||| ::|||
311 CCTTCGACACCATGCTCACAATTCGATTCGTATGAAGCCGGTAGTCCC 360
    ::||| ::||| ::||| ::||| ::||| ::||| ::|||
939 .....LeuLeuGlnHisValProProThrSerProProLysSerPr 952
    ::||| ::||| ::||| ::||| ::||| ::||| ::|||
361 GTTGCAGATCA.....GCCTTACCGCA..... 385
    ::||| ::||| ::||| ::||| ::||| ::||| ::|||
952 oValSerSphThrLeuSerGlyAlaLeuThrSerProProLysGlyP 969
    ::||| ::||| ::||| ::||| ::||| ::||| ::|||
386 .....TCCATTGGAGCGGATACGACACCATCCG.....CCGACG 421
    ::||| ::||| ::||| ::||| ::||| ::||| ::|||
969 rCProlaThrLeuAlaGluThrProThrTyProLysLysSerProLys 985
    ::||| ::||| ::||| ::||| ::||| ::||| ::|||
422 GCTATGACGGCGCAGCGCGGCTATCCGCTCCCAAGAGCGCGAGG 471
    ::||| ::||| ::||| ::||| ::||| ::||| ::|||
986 ProAlaAlaSerLysThrProAlaThrProSerProGluGly.... 1000
    ::||| ::||| ::||| ::||| ::||| ::||| ::|||
472 GATTATACAGCTACGACATATAAGCGCTGCCAAATATCCGCTCA 521
    ::||| ::||| ::||| ::||| ::||| ::||| ::|||
1001 .....ValThrAlaValPro...LeuGluThrProProLysSerLysVa 1015
    ::||| ::||| ::||| ::||| ::||| ::||| ::|||
522 CCTGACCGAACAACCGACGACCGGAGCTGTGCGACCTTTCACCA 571
    ::||| ::||| ::||| ::||| ::||| ::||| ::|||
1015 1a...ProLysThrAlaAlaProLysGluSerSerAlaThrSerSer... 1029
    ::||| ::||| ::||| ::||| ::||| ::||| ::|||
572 ATACCGGTGATGCTGACGCAAGAGTAGCGACGATTCAAACGCGCC 621
    ::||| ::||| ::||| ::||| ::||| ::||| ::|||
1030 .....SerLysArgAl 1033
    ::||| ::||| ::||| ::||| ::||| ::||| ::|||
622 ACCCGATACAGCC.....CCGAGCTGACAGATCGGGCA 656
    ::||| ::||| ::||| ::||| ::||| ::||| ::|||
1033 aProLysThrAlaValSerLysGluThrProSerLysGlyValThrAlaV 1050
    ::||| ::||| ::||| ::||| ::||| ::||| ::|||
```

```
657 TGGCGCCGAGCTTCAACGGCACTGCAGATATGCTCAAAACATCATCG 706
    ::||| ::||| ::||| ::||| ::||| ::||| ::|||
1050 aLProLeuGluThrLeuProLeuLysGluThrSerLysSerAlaThr 1066
    ::||| ::||| ::||| ::||| ::||| ::||| ::|||
707 GCGGCGAGGAGAAATTTGCGCGCAGCGCATGCGCTGACGGTATTAAC 756
    ::||| ::||| ::||| ::||| ::||| ::||| ::|||
1067 ProGlyLysSerAlaSerProLysArgSerProLysThrAlaGlu 1083
    ::||| ::||| ::||| ::||| ::||| ::||| ::|||
757 GAAGGCTCAACA.....TGGCTGTATGACGCGCTGGGTC 794
    ::||| ::||| ::||| ::||| ::||| ::||| ::|||
1083 yProLysGluThrProProGlyGlyValThrAlaValProProGluL 1100
    ::||| ::||| ::||| ::||| ::||| ::||| ::|||
795 GCTTTCACCGCAAAACAAGATGCGCGCATCAAGATTTGGCAGATATGG 844
    ::||| ::||| ::||| ::||| ::||| ::||| ::|||
1100 erLeuProProLysGluThrProGlnAsnAlaThrProAsnGluSerLeu 1116
    ::||| ::||| ::||| ::||| ::||| ::||| ::|||
845 .....CGCACTCAAAAGCT..... 859
    ::||| ::||| ::||| ::||| ::||| ::||| ::|||
1117 AlaAlaSerSerGlnLysArgSerProLysThrSerValProLysGluTh 1133
    ::||| ::||| ::||| ::||| ::||| ::||| ::|||
860 .....ATGCCGACGACGCCAT 875
    ::||| ::||| ::||| ::||| ::||| ::||| ::|||
1133 rProProGlyGlyValThrAlaMetProLeuGluThrProSerAlaProG 1150
    ::||| ::||| ::||| ::||| ::||| ::||| ::|||
876 CCGCATTTGGCGATCCCAAAACCCCA..... 901
    ::||| ::||| ::||| ::||| ::||| ::||| ::|||
1150 InLysAlaProLysThrAlaValAlaProLysGluThrProThrProGlu 1166
    ::||| ::||| ::||| ::||| ::||| ::||| ::|||
902 .....ATGCCGACGACGCCATAGAA 921
    ::||| ::||| ::||| ::||| ::||| ::||| ::|||
1167 AlaValThrIleLeuAlaGlySerProLeuSerProLysAlaSerL 1183
    ::||| ::||| ::||| ::||| ::||| ::||| ::|||
922 GCCGTCAGCAATATCTTATGGCAGTATCCCGTCAAAAGGATTTGAGC 971
    ::||| ::||| ::||| ::||| ::||| ::||| ::|||
1183 sThrAlaAlaProLysGluAlaProAlaThrProSerValGlyVal.... 1198
    ::||| ::||| ::||| ::||| ::||| ::||| ::|||
972 TGTTCGGGGAATAACGCTTGGCGCATTCAGCGCACATCTGTCAACG 1021
    ::||| ::||| ::||| ::||| ::||| ::||| ::|||
1199 .....IleAlaValSerGlyLysLeuSerProSer 1208
    ::||| ::||| ::||| ::||| ::||| ::||| ::|||
1022 GCTCGAGATGGCGAGATGCGATTCGCCAAGGAAATCCCGCTCAGC 1071
    ::||| ::||| ::||| ::||| ::||| ::||| ::|||
1209 ProLysThrSerLysThrAlaAlaProLysGluAsn.....SerAl 1223
    ::||| ::||| ::||| ::||| ::||| ::||| ::|||
1072 GAAATTTTCCGATGCGCATACGCCAATACCCGCTTACCATTC 1121
    ::||| ::||| ::||| ::||| ::||| ::||| ::|||
1223 aThrLeuProProLysArgSerProLysThrAlaAlaProLysGluThrP 1240
    ::||| ::||| ::||| ::||| ::||| ::||| ::|||
1122 CCGAAATATCCGTTCAAACTTGGAGCAGCGCTTACGCAAG...AAAACA 1168
    ::||| ::||| ::||| ::||| ::||| ::||| ::|||
1240 roAlaThrSerSerGluGlyValThrAlaValProSerGluThrPro 1256
    ::||| ::||| ::||| ::||| ::||| ::||| ::|||
1169 TCACCTCCTCAACCGTCCG..... 1188
    ::||| ::||| ::||| ::||| ::||| ::||| ::|||
1257 SerProProThrProAlaSerLysGlyValProValThrLeuThrPro 1273
    ::||| ::||| ::||| ::||| ::||| ::||| ::|||
1188 ..... 1188
    ::||| ::||| ::||| ::||| ::||| ::||| ::|||
1273 sGlyAlaProAsnAlaLeuAlaGluSerProAlaSerProLysValP 1290
    ::||| ::||| ::||| ::||| ::||| ::||| ::|||
1189 .....CCGTCAACGGAAG 1203
    ::||| ::||| ::||| ::||| ::||| ::||| ::|||
1290 rOlyThrAlaAlaProGluGluThrSerThrThrProSerProGluLys 1306
    ::||| ::||| ::||| ::||| ::||| ::||| ::|||
1204 AATGTGAACAGCGCA.....AACAA 1223
    ::||| ::||| ::||| ::||| ::||| ::||| ::|||
1307 IlProLysValAlaGlyProLysGluAlaSerAlaThrProProSerLy 1323
    ::||| ::||| ::||| ::||| ::||| ::||| ::|||
1224 ACGCCACCCGAAGCAAAAGTCCGTTGACGCTTAAAGGTTCCGAT 1273
    ::||| ::||| ::||| ::||| ::||| ::||| ::|||
1323 sLysThrProLysThrAlaValPro..... 1331
    ::||| ::||| ::||| ::||| ::||| ::||| ::|||
1274 TTGAAAAAGACGTAAATATACGATACGAAATTAATACCGCTGTAACA... 1320
    ::||| ::||| ::||| ::||| ::||| ::||| ::|||
```

```

1332 .....LysGluThrSerAlaProSerGluGlyValThrAlaValProLeu 1346
1331 .....CAAGTAAATCCATAGA 1337
1347 GuileProProSerProAlaGlyAlaProLysThrAlaAlaProLysGlu 1363
1338 TGACCCGCTTTTAACTCAAAAGGTTCTGCGATCGGCTCATCTTGGT 1387
1363 .....LysGluThrAlaProSerProGluGlyAlaThr..... 1373
1388 CTATACGCCCAATTCATATACGAAATTCACAGCAGCAAGTGAATC 1437
1374 .....ThrAlaProValGlnIleProProSerProAlaGlySerLys 1388
1438 AGATATATCCACCTAAATTAATCTCCCTGACGACCGCTCAACAAAG 1487
1389 Lys.....AlaGlySerLysGluThrProThrProSerProGluGlu 1403
1488 A 1488
1403 Y 1403

```

seq_name: p1r2:A43932

```

seq_documentation_block:
  mucin 2 precursor, intestinal - human (fragments)
  N/Alternate names: mucin SMUC-41
  C/Species: Homo sapiens (man)
  C/Date: 10-Mar-1993 #sequence revision 12-Apr-1996 #ext_change 05-Nov-1999
  C/Accession: A49963; A45106; B45106; A43932; B03322; A61257; P00329
  R/Gum JR.: J.R.; Hicks, J.W.; Toribara, N.W.; Siddiki, B.; Kim, Y.S.
  J. Biol. Chem. 269, 2440-2446, 1994
  A>Title: Molecular cloning of human intestinal mucin (MUC2) cDNA. Identification of the
  A:Reference number: A49963; MUID:94132002
  A:Accession: A49963
  A:Molecule type: mRNA
  A:Residues: 1639 <GU1>
  A:Cross-references: GB:I21998
  R/Gum JR.: J.R.; Hicks, J.W.; Toribara, N.W.; Rothe, E.M.; Lagace, R.E.; Kim, Y.S.
  J. Biol. Chem. 267, 21375-21383, 1992
  A>Title: The human MUC2 intestinal mucin has cysteine-rich subdomains located both upstr
  A:Reference number: A45106; MUID:93016075
  A:Accession: A45106
  A>Status: not compared with conceptual translation
  A:Molecule type: mRNA
  A:Residues: 626-1895 <GU2>
  A:Cross-references: GB:M94131; NID:9186395; PIDN:AAA59163.1; PID:9186396
  A>Note: sequence extracted from NCBI backbone (NCBIP:110706)
  A:Accession: B45106
  A>Status: not compared with conceptual translation
  A:Molecule type: mRNA
  A:Residues: 2037-3020 <GU3>
  A:Cross-references: GB:M94132; NID:9186397; PIDN:AAA59164.1; PID:9186398
  A:Experimental source: colon
  A>Note: sequence extracted from NCBI backbone (NCBIP:116698)
  R/Toribara, N.W.; Gum JR.: J.R.; Culhane, P.J.; Lagace, R.E.; Hicks, J.W.; Petersen, G.M.
  J. Clin. Invest. 88, 1005-1013, 1991
  A>Title: MUC-2 human small intestinal mucin gene structure. Repeated arrays and polymorp
  A:Reference number: A43932; MUID:91358717
  A:Accession: A43932
  A:Molecule type: DNA
  A:Residues: 1343-1350, 'L', 1352-1411, 'S', 1413-1448, 'P', 1450-1503, 'T', 1505-1915 <TOR>
  A:Cross-references: GB:M74027; NID:9188863; PIDN:AAA59875.1; PID:9188864
  A>Note: sequence inconsistent with the nucleotide translation
  R/Gum JR.: J.R.; Hicks, J.W.; Toribara, N.W.; Lampert, D.T.A.; Kim, Y.S.
  J. Biol. Chem. 264, 6480-6487, 1989
  A>Title: Molecular cloning of human intestinal mucin cDNAs. Sequence analysis and eviden
  A:Reference number: A33532; MUID:89197956
  A:Accession: B33532
  A:Molecule type: mRNA
  A:Residues: 1916-2193 <GU4>

```

```

A:Cross-references: GB:M22405; NID:9188873; PIDN:AAA6334.1; PID:9188874
A:Experimental source: intestine
R/Jany, B.H.; Gallup, M.W.; Yan, P.S.; Gum, J.R.; Kim, Y.S.; Basbaum, C.B.
J. Clin. Invest. 87, 77-82, 1991
A>Title: Human bronchus and intestine express the same mucin gene.
A:Reference number: A61257; MUID:91086481
A:Accession: A61257
A>Status: not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 'T', 1925-1948, 'TTS', 1952-1954 <JAN>
A:Experimental source: bronchus
R/Xu, G.; Huan, L.; Khatri, I.; Sajjan, U.S.; McCooll, D.; Wang, D.; Jones, C.; Forstn
Biochem. Biophys. Res. Commun. 183, 821-828, 1992
A>Title: Human intestinal mucin-like protein (MLP) is homologous with rat MLP in the
A:Reference number: P00328; MUID:92198477
A:Accession: P00328
A:Molecule type: mRNA
A:Residues: 2328-2468 <XUG>
A:Cross-references: GB:M86523
A:Experimental source: small intestine
A:Accession: P00329
A:Molecule type: protein
A:Residues: 2328-2342, 'K', 2344-2354 <XUG1>
C/Genetics:
A:Gene: GDB:MUC2
A:Cross-references: GDB:120203; OMIM:158370
A:Map position: 11p15.5-11p15.5
C/Superfamily: von Willebrand factor; von Willebrand factor type A repeat homology; v
C/Keywords: glycoprotein; intestine; tandem repeat
F:2766-2834/Domain: von Willebrand factor type C repeat homology <WVC>

```

```

alignment_scores:
  Quality: 153.50      Length: 315
  Ratio: 1.104        Gaps: 13
  Percent Similarity: 94.127      Percent Identity: 24.444

```

alignment_block:

US-09-303-518D-465 x A43932 ..

```

Align seg 1/1 to: A43932 from: 1 to: 3020
269 ACATTCGCGCTTTTCGATCAGCGGAGCAAGTCATCCCTTCGAC 318
1520 ThrThrThrProSerProThrThrThrThrThrThrProProth 1536
319 AACCATGCTCATTCGATCTGATGAAGCGGAGTCCCGTTCGACG 368
1536 ThrThrThrProSer...ProThrThrThrProThrProthProth 1552
369 ATTCAGCTTTTACCGCATTCGATTCGATGAGCATGACACATCCGCGC 418
1552 erthr.....ThrThrLeuProProThrThrThrThrProProPro 1566
419 ACGGCTATAGAGGCGCACAGGCGGCGCTATCCGCTCCCAAGGCGG 468
1567 ThrThrThrThrThrProProProThrThrThrProSerPro..... 1580
469 AGGAGATATACAGCATGACGATGAAAGCGCTTCCCAAAATATCCGCT 518
1581 .....ProThrThrThrThrProProProProthThrThrThr 1594
519 CAACGTCAGCGACAAACGCGACGCGACGAGCGCTGTCGACGCTTCC 568
1594 hrThrProProProThrThrThrThrProSerProProThrThrThr 1610
569 ACATATCCGCTGATGCTGACGAGGAGTGGCGAGGATCAACGC 618
1611 ThrProPro.....ProThrThrThrProSe 1619
619 GCACCCGATACGCGCGAGCTGGACAGATGCGGCAATGCGCGGAGC 668
1619 rProProThrThrThrPro.....lleThrProProthrs 1631

```



```

802 ACCGGAACCAAGATGGCGGCATCAACAGATTGGCAGATVGGCGCACT 851
252 abProthr.....SerAlaThrSerValGlnProSerSerSerS 267
853 CAAGACATATGCCGAG.....CAGCATCCCGGATTGGGCG 889
267 eThrProleProSerThrThrSerValGlnProSerSerSerSer 283
890 TCCAAAACCCA.....ATCCCGCAAGAGCATGAAAGC 924
284 AlaProThrThrSerAlaThrSerValGlnProSerSerSerSer 300
925 G.....TCACCATATCTTTACGGCAGTATCCCGTCAAGGAT 965
300 OProleSerSerThrIleSerValGlnProSerSerSer..... 314
966 TGGAGCTGTTCGGGGAATAACGCGTTGGCGGCATCAGGCAATCTCG 1015
315 .....SerSerProThrThrSerThrThrSerThrSerValGlnProSer 327
1016 TCACAGCGCTGCAGATGGCGGAGATCG.....CAVTGCGC 1050
328 SerSerGlySerAlaProThrThrSerAlaThrSerValGlnProSerSe 344
1051 AAAGGAAATCCGGCCGTCACCGCAATTTGCCGATGGCGCATACGCCAA 1100
344 rSerSerSerProProleSerThrIleSerValGlnProSerSers 361
1101 ATACCGCTCCCTTACCATTCCGAAATATCCGTTCAAACTGGAGCAGC 1150
361 eSerSerSerProThrThrSerThrThrSerValGlnProSerSer 377
1151 GTTACGGCAAGAAACATCAC.....TCCTC 1178
378 GlySerAlaProThrThrSerAlaThrSerValGlnProSerSerSe 394
1179 AACCGTCGCCCGTCAACGAAAGATGTGAACGTGGCAAAACAAACCC 1228
394 rSerValProThrThrSerAlaThrSerValInrSerSerSer.... 409
1229 ACCCGAAGACCAAGTGCCTTTGACGCGTAAAGGGTTCCGAATTTGAA 1278
410 ..SerSerThrProleProThrThrThrSerValGlnProSerSer 425
1279 AAGAGCTAAATAATACGAT.....ACGAGATATATACCGCTGACCA 1322
426 SerSerValProThrThrSerAlaThrSerValGlnThrSerSerSe 442
1323 AGTGAATCCTATAGATGAACCCGTTTAAATCCAAAGTCTCTCGGAT 1372
442 rSerThrProleProSerThrThrSerValGlnProSerSerSers 459
1373 CGGCTATCTTGTCGTATACTGCC 1398
459 eAlaProThrThrSerAlaThrSer 467

```

```

seq_name: p1r2:T45134
seq_documentation_block:
  hypothetical protein [imported] - Microbacterium ammoniaphilum (fragment)
  C:Species: Microbacterium ammoniaphilum
  C:Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 21-Jul-2000
  C:Accession: T45134
  R:Striebel, H.M.; Seeber, S.; Jarsch, M.; Kessler, C.
  Gene 172, 41-46, 1996
  A:Title: Cloning and characterization of the MamI restriction-modification system from
  A:Reference number: 422923; MUID:96257230
  A:Accession: T45134
  A:Status: preliminary; translated from GB/EMBL/DBJ
  A:Molecule type: DNA
  A:Residues: 1-529 <STR>
  A:Cross-references: EMBL:X79027; NID:9984667; PIDN:OAA55649.1; PID:q1679831

```

A: Experimental source: ATCC 15354

alignment_scores:
 Quality: 153.50 Length: 427
 Ratio: 0.872 Gaps: 22
 Percent Similarity: 41.218 Percent Identity: 25.293

alignment_block:
 US-09-303-518D-465 x T45134 ..

Align seg 1/1 to: T45134 from: 1 to: 529

```

100 CGCGCAGTT.....CTGACCGTCAGCATTTTCGACCCGACGGGAATA 143
    ||||| ||||| ||||| ||||| |||||
78 ArgGlnValHisProLeuGlnArgProHisGlnInuPro.....90
144 CCACGATTTCGGCAGCAGGGGGGAACCTGC.....CGACCG 180
    ||||| ||||| ||||| ||||| |||||
91 .....GlyAlaIatrgLylGluLeuProValArgAspAspArgAlaG 104
181 AGCGGTCAATTCGG.....ATTGGAAATCATCAAAACCA 215
    ||||| ||||| ||||| ||||| |||||
104 IuArgGlyArgArgGluProAlaItrProAlaIaIatrgAspAlaIatrg 120
216 TCAGTTGGGGCAACCTGTTCACTCCAGCA...GGCGGCATTAAAGAAATA 262
    ||||| ||||| ||||| ||||| |||||
121 HisLeuAlaIrgLeuGlyAlaHisProAlaIatrgArgGlyValLeuItrg 137
263 TCGGCTACATTGTCCGGTTTCCGATCAGCGGCA...CGAAGTCATTCC 309
    ||||| ||||| ||||| ||||| |||||
137 sArgItrHis.....ArgAlaIatrgGlyGluIatrgGlyArgGlyP 150
310 CCCTTCGACGACCACTGCGCTCACTCCGATTCCGATTGATGAACCGGTAGTCC 359
    ||||| ||||| ||||| ||||| |||||
150 ToArgGlnInL..... 153
360 CGTTGACGAGATTTCACGCTTACCGCATTCATGGGACGGATACGACACC 409
    ||||| ||||| ||||| ||||| |||||
154 .....ValProIatrgGlnHisProIatrgGlyIatrgArgAspArgAl 166
410 ATCCCGCCGACGCGCTATGACGGGGCACAGGGCGCGGCTATCCCGCTCCG 459
    ||||| ||||| ||||| ||||| |||||
166 agLIatrgProGlyLeu.....HisatrgAtrgAtrgArgAlaIa 179
460 AAAGCGCGAGGGGATATATACACTGACGACATTAAGGCGTTGCCCAAA 509
    ||||| ||||| ||||| ||||| |||||
179 rGatIatrgGlnGlyIatrgProGlnValatrgHis.....189
510 TATCCGCGCTCAACCTGACCGCAACCGACGACCGGACACAGCGCTTGTG 559
    ||||| ||||| ||||| ||||| |||||
190 .....GlyAspAspGlnHisIatrgAlaIatrgAspProIatrg 200
560 ACGGTTTCCCAATTACCGGTAATGCTATGCTACGCCAAGAGTACGACGGA 609
    ||||| ||||| ||||| ||||| |||||
200 gProIatrgAspProIatrgAlaGlyHisHisProLeuIatrgGlnGlyGlnG 217
610 TTCAAGCGCCGACCGGATACAGACCGCCGAGCGTGCAGATGCGG..... 653
    ||||| ||||| ||||| ||||| |||||
217 IyGlnGlnIaIatrgProLeuIatrgAlaIatrgGlyGlyGlnGlyIatrg 233
653 ..... 653
234 GLyProIatrgLIatrgHisProAlaLeuGlyAspGlyIatrgAspProIatrg 250
654 .....CATGCGCGCGGACGCTTTCAAACGCGCATGCGATATGCGTCAAAA 699
    ||||| ||||| ||||| ||||| |||||
250 aATatrgGlnGlyAlaIatrgAlaIatrgAlaIatrgGlyIatrgHisLeuIatrg 267
700 ATATGCGCGCGCGGACGACAAAT...TTCGCG.....CGC 731
    ||||| ||||| ||||| ||||| |||||
267 IatrgIatrgIatrgGlyIatrgAlaIatrgGlyCysAspIatrgAlaGlnGlyIatrg 283

```

```

732 AGCGATGCGCGAGGTATTAAGCGAAGCGTCAACATTGCTGTATGC 781
    ||| |||||
284 ArgAlaArgArgAlaGly..... 289
782 AGCGTGGGTGCTTCCACCGAAACAAAT.....GGCGGCAT 824
    ||| |||||
290 .....GlyGlyGlyLeuProArgArgGlyAspArgValArgAlaHis 304
825 .....CAACGATTGGCAGATATGGCGCACTCAAGACTATGCCGCG 869
    ||| |||||
304 rgrProArgProArgGlyArgGlyAlaAla..ArgLeuAspArgAla 319
870 AGCGATCGCGATGGCGAGTCCAAACCCCAATGCCGCAAGGCATAG 919
    ||| |||||
320 GlyGlyAlaGlyAlaArgProAlaArgProArg...ArgLeuArg..... 333
920 AAGCGGTACGAATATCTTAAACGCGACATGCCGTCAAAGGGATGGA 969
    ||| |||||
334 .....HisProArgProAlaAspLeu 341
970 GCTGTGCGGGAATAACGGCTTGGCGG...CATCAGCGCATCTGT 1016
    ||| |||||
341 rGHisGlyGlyProGlnGlyAlaAlaArgLeuAspHisProGlnGlyLeu 357
1017 CAAGCGGTGCGAGATGGCGCGATCGCATTCGCAAGGGAATCCGCG 1066
    ||| |||||
358 GlnGlyAlaAlaGlySerArg..... 364
1067 TCAGGACAAATTGCGCATGCGCATACGCCAAATACCGCTCCCTTAC 1116
    ||| |||||
364 ..... 364
1117 CATTCGCCGAATATCCGTTCAACTGGAGACGCTTACGCAAGAAAA 1166
    ||| |||||
365 .....ArgHisProHisArgLeuArgAlaArgLeuHisGlnGlyArg 378
1167 CATCACTCTCAACCGTCGCGCGCTCAACGCAAGAAATGTGAACCTGG 1216
    ||| |||||
379 GlyAspLeuLeuArgArgProArgArgAspArg.....LeuG1 391
1217 CAACAAACGCGCACCGCAAGACCAAAAGTCC 1247
    ||| |||||
391 YArgArgGlyProArgGlnGlyAla 401

seq_name: pir2:S35047
seq_documentation_block:
mucin JUL7 - human
C:Species: Homo sapiens (man)
C:Date: 10-Dec-1993 #sequence_revision 10-Nov-1995 #text_change 08-Oct-1999
C:Accession: S35047
R:Dufose, J.; Porchet, N.; Audle, J.P.; Guyonnet Duperrat, V.; Laine, A.; van-Seuningen,
Biochem. J. 293, 329-337, 1993
A:Title: Degenrate 87-base-pair tandem repeats create hydrophilic/hydrophobic alternati
A:Reference number: S35047; MUID:9343838
A:Accession: S35047
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-543 <DUP>
A:Cross-references: EMBL:X74370; NID:9407081; PID:CA52408.1; PID:9407082
A:Note: the authors translated the codon AAA for residue 63 as Gln and CGG for residue 3
A:Note: the amino acid sequence from fig. 2 is inconsistent with the nucleotide sequence

```

```

alignment_scores:
  Quality: 149.50      Length: 395
  Ratio: 0.808        Gaps: 16
  Percent Similarity: 46.835      Percent Identity: 22.025

```

```

alignment_block:
US-09-303-518D-465 x S35047 ..
Align seg 1/1 to: S35047 from: 1 to: 543

```

```

128 AACCCGAGGGGAATACACCTATTCCGACAGGCGGAATGCGGAG 177
    ||| |||||
44 SerProLeuAspSerThrAlaThrProSerThrProGlyThrAlaPro 60
178 GCGACGCGTCATATCGATTGGGAACATCAAGACCATGTTGGCGAA 227
    ||| |||||
60 ProGlyValLeuThrSerProAlaThrThrProThrAlaThrSerSer 77
228 CCGTTCATCCAGCAGCGCGCATTAAGAAATATCGGTCACATTCGCC 277
    ||| |||||
77 yAlaThrSerSerSerPro.....ArgThrAlaThrThrLeuPro 91
278 GCTTTCCGATCAGCGGACGAAGTCATTCCTCCCTTGACACACATGCC 327
    ||| |||||
92 ValLeuThrSerThrAlaThrLysSerThrAlaThrSerValThr.... 106
328 TCACATTCGATTCGATGTAAGCCGTAAGTCCCGTTGACGAGATTACGCT 377
    ||| |||||
107 .....ProIleProSerSerThrLeuGlyT 115
378 TTACCGCATCCATTGGGACGATACGACACCATCCCGCGCGGCTATG 427
    ||| |||||
115 hrThrGlyThrLeuProGlnGlnThrThrThrProValAlaThrMetSer 131
428 ACGGCGCACAGGCGCGCGCTATCCCGTCCCAAGGCGCGAGATATA 477
    ||| |||||
132 ThrIleHisProSerSerThr.....ProGlnThrThrHisThrSe 145
478 TCACGTCACGACATAAAGCGCTGCCCAAAATATCCGCTCAACCTGAC 527
    ||| |||||
145 ThrValLeuThrThrLysAla..... 152
528 CGACACCGGACGACCGGACAAAGCGCTTGCACGCTTTCACAAATACCG 577
    ||| |||||
153 ..ThrThrThrArgAlaThrThrSerThrSerThrProSerSerThrPro 168
578 ...GTAGATGCTGACGCAAGAGTAGCGGAGATTCAAAGCGCGCAC 624
    ||| |||||
169 GlyThrThrThrIleLeuThrGlnLeuThrThrAlaAlaThrThrThrAl 185
625 CGATACAGCGCGCGAGTGGACAGATCGGCAATGCCCGCGAAG.....C 668
    ||| |||||
185 aglyThrGlyProThrAlaThrProSerSerThrProGlyThrThrTrp1 202
669 TTTCACGCGCAGTCAGATATGTCATAAATCATCATCGCGCGCGAGAG 718
    ||| |||||
202 leuThrThrGlnLeuThrThrThrAlaThrThrThrAlaSerThrGlySer 218
719 AATTGTCGCGCGAGGCGATGCCGTGACGGGTAAAGCGAAGCTCAAC 768
    ||| |||||
219 ThrAlaThrLeuSerSerThrPro..... 226
769 ATTGCTGTATGCAAGCGCTTGGGTGCTTTCACCGCAAAACAAATGGC 818
    ||| |||||
227 .....GlyThrThrThrIleLeuThrGlnProSerThr..... 237
819 GCGCATCAAGATTGGCAGATATGGCGCAACCAAGACTATGCGCGAG 868
    ||| |||||
238 ..ThrAlaThrVal.....ThrValProThr 245
869 CAGCATCCGCGATTTGG.....CAGTCCAAACCCCAATGCCGCA 909
    ||| |||||
246 GlySerThrAlaThrAlaSerSerThrGlnAlaThrAlaGlyThrProH1 262
910 CAAGCATACAAAGCCGTCAGCAATATCTTTACGACATCATCCCG... 955
    ||| |||||
262 sValSerThrThrAlaThrThrProThrValThrSerSerLysProLeu 279
956 .....TCAAGGGAATTGGAGCTTCGCGAAATATACGCTTGCGCGCA 1000
    ||| |||||
279 euleuSerProGlyLeuGlnLeuProSerAlaLeuArgSerThrAlaThr 295

```



```

715 ProAlaAsnIleProCysIleValProThrProAlaProAlaProAl 731
62 ATGCACAGCGCTCAGATTGGCAACGATCTTTATCCGGCAGGTTCC 111
731 a.....HisPheSerGlnProValSerSerG 740
112 GACCGTCAGCAT..... 124
740 InArGValAlaGlnGlnGlnGlnAsnThrLeuGlnIleValAlaLeuAsnAsp 756
125 .....TCGAACCGGAGGGAATGCCACTATTC 153
757 GluLeuLysGlyAsnLeuAsnLysArgProThrGlyThrThrAlaProPr 773
154 GGCAGCAGGGGGGAACTTGGCAGCGGTCATATCGATTGGGAAA 203
773 oSerAsnGlyPheAsnAlaProArgAlaAspValAlaProValGln. 789
204 CATCAAGACCATCAGTTGGCAACCTGTTCATCCAGCAGG...CGGCCA 250
790 .....ArgProIleSerSerAlaSerIleProAlaLeuGlnProGlnPro 804
251 TTAAGGAATATCGGCTACATGTCGGCTTTCCGATCACGGGAGCA 300
805 ILeGlnHisIleGlnLysProIleGlnProGlnIleVal..... 817
301 GTCCATTCGCCCTTGACAAACCATGCTCATCTCGATTCTGATGAGC 350
818 ArgIleProProSerThrAlaProValGlnLysProValGlnValSerA 834
351 CG.....GTAGTCCCGTTGACGGATTCAGCCTTTACGCCA 385
834 LaProThrHisSerAsnValAlaProThrThrSerGlnAlaSerAla 850
386 TCACATTGGAGCGATACGACACCATCCCGCG..... 418
851 AspAla.....ArgAsnProLeuProProLysThrSerProProVa 864
419 .....ACGGCTATGACGGGCCACAGGGCGGGCTATTC 452
864 IserAsnThrProIleThrValAlaProValHisAlaProThrThrS 881
453 CGCTCCCAAGCGCGAGGATATAT.....ACAGTACGACATTA 493
881 eAlaIaProSerThrSerValValThrArgArgProThrSerThrAla 897
494 AAG..... 496
898 GlnMetSerAspLugIuaArgSerArgIleAlaMetAspIleSerSe 914
497 GCGTTGGCCCAAAATATCCGCC.....TCAACCTGACCGCAGACCG 536
914 rAlaLeuProAlaProSerAlaLeuLeuLysGlySerAsnSerThrS 931
537 CAGCAGCGCAACGCGTTGACCGTTTCCACATACCGGTAGTATGC 586
931 eLysProSerAlaAlaValSerThrAlaSerSerValPro..... 944
587 TGCAGCAAGAGTAGGAGCGGATTCACAGCGGCCACCGCATACAGCCC 636
945 .....SerThrAlaAlaArgAsnProValGlnThrArgTr 956
637 GAGCTGAGAGATGCGCAATGCGCGGAGAGCTTTCACAGCGCATGCGACA 686
956 oSerGlnProHisValThrMetProLysLysSerSerGlnProIleL 973
687 TATCGTCA.....AAACATCATCGCGCGGCGAGGAG 718
973 eSerSerGluValLeuGlnProThrArgLeuProSerThrAlaThrSer 989
719 AAATTTGCGCGAGCGAGATCCGTCAGGGGTAAAGCGAAGGCTCAAC 768
990 GlnAlaLysProValThrGlnProIleArgHisProSerProProValAl 1006

```

```

769 ATTCCTGTATGACAGCGCTTGGCTCTGCTTCCACGAAACAGATGCG 818
1006 AtThrValIleProThrAlaValAlaAspLysLysProValSerGlnSng 1023
819 GCGCATCAACGATT.....TGCAGATATGGCGCAACGAACTATG 862
1023 InGlySerAsnValProLeuPheAsnIleThrAsnSerSerAsnGlyTyr 1039
863 CCGCAGCAGCATCCCGCATGGCAGTCCCAAAACCCCATGCGCGACAA 912
1040 ProGln.....LeuAsnGlyTyrProAsnTyr 1048
913 GGCATAGAAGCGCTCAGCATATCTTTACGCGAGTCAATCCCGCTCAAG 962
1048 rGlyAsnGlyPheGlnAlaTyrGlyTyrGlyMetAsnTyrHisGln.Gl 1064
963 GATTGAGCTGTTCGGGA.....AAATACGCTTGGCGCGCATCAGC 1006
1007 CACATCTGTCAAGCGGTGCGAGATGCGGAGATGSCATTCGCGAAAGG 1056
1078 .....MetGlyGlnLeuAlaLeuThrHisAsn 1086
1057 AAATCGCGCGTCAGCAGACAAATTTCGCAATGCGCATCGCCAAATACC 1106
1087 .....AlaValThrSerLeuPr 1092
1107 GTCCCTTACATTCGCCGAATATCCGTTCAACCTGGAGCAGCCTTACG 1156
1092 oProLeuValProSerGluAsnArgPheSerGlyThrAlaGlnProLeuG 1109
1157 GCAAAGAAACATC..... 1170
1109 LysIleSerAspIleMetGluPheLeuGlyThrGlnGlnArgGlnAlaGly 1125
1171 ...ACCTCTCAACCGCTGGCGCGTCGAACGGAAG..... 1203
1126 SerSerSerArgAlaValProProAlaSerAlaSerThrSerAlaLase 1142
1204 .....AATGTGAACCTGGCAACAACGCCACCGGAAGACCAAG 1243
1142 rGlyIleThrAspLeuSerMetAlaAspLysMetGluValLeuTyrArg. 1158
1244 TGCCGTTTACGCGTAAAGGTTTCCGATTTGAAAGACGTAATAATC 1293
1159 .....GlnAlaAspThrHis 1164
1294 GATACGAGAAATTAATACCGCTGTACCAACAAGTGAAT 1329
1165 LysGlyAsnGlyAspThrMetValSerGlnCysAsn 1176

seq_name: p1r1:S48478
seq documentation block:
glucan 1,4-alpha-glucosidase (EC 3.2.1.3) - yeast (Saccharomyces cerevisiae)
N:Alternate names: extracellular glucosylase; mncin-like protein MNC1; protein Y1R01
C:Species: Saccharomyces cerevisiae
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 12-Nov-1999
C:Accession: S48478; A26877; B26877; S27281; J06123
R:Rowley, K.
submitted to the EMBL Data Library, October 1994
A:Reference number: S48478
A:Accession: S48478
A:Molecule type: DNA
A:Residues: 1-1367 <R0>
A:Cross-references: GB:Z47047; EMBL:Z38061; NID:g603997; PID:g763364; GSPDB:GN00009;
R:Yamashita, I.; Nakamura, M.; Fukui, S.
J. Bacteriol. 169, 2142-2149, 1987
A:Title: Gene fusion is a possible mechanism underlying the evolution of STR1.
A:Reference number: A91831; MID:87194600
A:Accession: A26877

```

A: Molecule type: DNA
 A: Residues: 1-242 <YAM>
 A: Cross-references: EMBL: M16164; NID: g172522; PIDN: AAA35014.1; PID: g172525
 A: Accession: B26877
 A: Molecule type: DNA
 A: Residues: 762-1331 <YAZ>
 A: Cross-references: EMBL: M16165; NID: g172523; PIDN: AAA35015.1; PID: g172526
 A: Pardo, J.M.; Ianez, E.; Zalacain, M.; Clarios, M.G.; Jimenez, A.
 FEBS Lett. 239, 179-184, 1988
 A: Title: Similar short elements in the 5' regions of the STR2 and SGA genes from Sacchar
 A: Reference number: S27281; MIMD: 89031230
 A: Accession: S27281
 A: Molecule type: DNA
 A: Residues: 1-31 <PAR>
 A: Cross-references: EMBL: X13857; NID: g4551; PIDN: CAA32069.1; PID: g4552
 A: Lambrechts, M.G.; Bauer, F.F.; Marmur, J.; Pretorius, I.S.
 Proc. Natl. Acad. Sci. U.S.A. 93, 8419-8424, 1996
 A: Title: MucL, a mucin-like protein that is regulated by Msi10, is critical for pseudohy
 A: Reference number: J06123; MIMD: 96523237
 A: Accession: J06123
 A: Status: nucleic acid sequence not shown
 A: Molecule type: DNA
 A: Residues: 1-1367 <LAM>
 A: Cross-references: GB: U030626; NID: g1304386; PIDN: AAC49609.1; PID: g1304387
 A: Gene: SGD: MUC1; STR2; MAL5; DEX2; SGD: S0001458
 A: Cross-references: MIPS: Y1R019c; SGD: S0001458
 A: Map position: 9R
 C: Superfamily: yeast glucan 1,4-alpha-glucosidase homolog; glucan 1,4-alpha-glucosidase
 F: 5-21/Domain: transmembrane #status predicted <TM1>
 F: 1350-1366/Domain: transmembrane #status predicted <TM2>

alignment_scores:
 Quality: 147.00 Length: 532
 Ratio: 0.546 Gaps: 18
 Percent Similarity: 50.564 Percent Identity: 20.865

alignment_block:
 US-09-303-518d-465 x S48478 ..

Align seg 1/1 to: S48478 from: 1 to: 1367

```

110 TCAGCCGTGACATTCGACACCGGAGGAAATACCACTATTCGGACG 159
    ::::| | | | | | | | | | | | | | | | | | | | | |
438 ThrThrgluserSerSerAlaProValThrSerThrThrgluser 454
    ::::| | | | | | | | | | | | | | | | | | | | | |
160 AGGGGGAACTTCCGACGCGACGCGTCATATCGGATTTGGAAACATACA 209
    ::::| | | | | | | | | | | | | | | | | | | | | |
454 rSerAlaProValProThrProSerSerThrThrgluserSer 471
    ::::| | | | | | | | | | | | | | | | | | | | | |
210 AACCCATCAGTTGGCAACCTGTTTCATCCAGAGGGCCATTAAAGAA 259
    ::::| | | | | | | | | | | | | | | | | | | | | |
471 lalProValThrSerSerThrThrgluserSerAlaProValProThr 487
    ::::| | | | | | | | | | | | | | | | | | | | | |
260 ATATCGGTACATG.....TCGGCTTTCGATACGAGGACGAA 300
    ::::| | | | | | | | | | | | | | | | | | | | | |
488 ProSerSerThrThrgluserSerSerAlaProValThrSerSer 504
    ::::| | | | | | | | | | | | | | | | | | | | | |
301 GTTCATTCCTCCCTCGACACCATGCTTCATTCGATTCGATGAGCAAGC 350
    ::::| | | | | | | | | | | | | | | | | | | | | |
504 rThrGluserSerSerAlaProValPro.....Thrp 515
    ::::| | | | | | | | | | | | | | | | | | | | | |
351 CGGTAGTCCCGTTGACGATTCAGCCTTACC.....GCA 385
    ::::| | | | | | | | | | | | | | | | | | | | | |
515 rSerSerSerThrThrgluserSerSerAlaProAlaProThrProSer 531
    ::::| | | | | | | | | | | | | | | | | | | | | |
386 TCACATTCGACGATACGACACATCCCGCGGCGATGAGCGGCGCA 435
    ::::| | | | | | | | | | | | | | | | | | | | | |
532 SerSerThrThrgluserSerSerAlaProValThrSerThrThrgl 548
    ::::| | | | | | | | | | | | | | | | | | | | | |
436 CAGGGCGGCGCT...ATCCGCTCCCAAGCGCGAGGATATATACAG 482

```

```

548 userSerSerAlaProValProThrProSer.....SerS 560
    ::::| | | | | | | | | | | | | | | | | | | | | |
483 CTACGACATTAAGCGCTTCCCAAAATATCCGCCATCAGCGACGACA 532
    ::::| | | | | | | | | | | | | | | | | | | | | |
560 erThrThrgluserSerSerThrProValThrSerThrThrgluser 576
    ::::| | | | | | | | | | | | | | | | | | | | | |
533 ACCGACGA.....CCGACACGCGCTTCCGCTTCCAC 570
    ::::| | | | | | | | | | | | | | | | | | | | | |
577 SerSerAlaProValProThrProSerSerThrThrgluserSer 593
    ::::| | | | | | | | | | | | | | | | | | | | | |
571 ATACCGGTAGTATGCTGACGACGAGAGTAGCGACGATTAACGCGC 620
    ::::| | | | | | | | | | | | | | | | | | | | | |
593 lalProValProThrProSerSerThrThrgluserSerSerAla 610
    ::::| | | | | | | | | | | | | | | | | | | | | |
621 CA...CCGATACAGCCCGAGCTGACGACATGCGCAATGCGCGAAG 667
    ::::| | | | | | | | | | | | | | | | | | | | | |
610 roAlaProThrProSerSerThrThrgluserSerSerAlaProVal 626
    ::::| | | | | | | | | | | | | | | | | | | | | |
668 CTTCACAGCGCCTGACATATGCTCA.....AAACATCATCGGC 708
    ::::| | | | | | | | | | | | | | | | | | | | | |
627 ThrSerSerThrThrgluserSerSerAlaProValProThrProSer 643
    ::::| | | | | | | | | | | | | | | | | | | | | |
709 GCGGACGAGAAATGTCGCGCAGCGCATGCCGTGACGAGATATACGA 758
    ::::| | | | | | | | | | | | | | | | | | | | | |
643 rSerThrThrgluserSerSerAlaProValProThr.....ProS 657
    ::::| | | | | | | | | | | | | | | | | | | | | |
759 AGGCTCAACATGCTGTTATGACGCGCTGCGTCTGTTCCACCGAA 808
    ::::| | | | | | | | | | | | | | | | | | | | | |
657 erSerSerThrThrgluserSerSerAlaProValProThrProSer 673
    ::::| | | | | | | | | | | | | | | | | | | | | |
809 ACAAGATGGCGGCATCAACG.....ATTGCGCATATGCGCACTC 852
    ::::| | | | | | | | | | | | | | | | | | | | | |
674 SerThrThrgluserSerSerAlaProValThrSerThrThrgluse 690
    ::::| | | | | | | | | | | | | | | | | | | | | |
853 AAAGATATGCGCAGACCATCCGATTCGCGCAAAACCCCAA 902
    ::::| | | | | | | | | | | | | | | | | | | | | |
690 rSerSerAlaProValThrSerSerThrThrgluserSerSerAlaPro 707
    ::::| | | | | | | | | | | | | | | | | | | | | |
903 TGCCG.....CACAGCATAGAACCGC 925
    ::::| | | | | | | | | | | | | | | | | | | | | |
707 alProThrProSerSerThrThrgluserSerSerAlaProValPro 723
    ::::| | | | | | | | | | | | | | | | | | | | | |
926 TCAGCATATCTTTACGCGACATCCCGTCACAAAGATGAGCTGT 975
    ::::| | | | | | | | | | | | | | | | | | | | | |
724 ThrProSerSerThrThrgluserSerSerAlaProValProThrPr 740
    ::::| | | | | | | | | | | | | | | | | | | | | |
976 CGGGAAATACGCGCTTGGCGGATCAGCGACATCTGTCAGCGGTC 1025
    ::::| | | | | | | | | | | | | | | | | | | | | |
740 oSerSerSerThrThrgluserSerSerAlaProValThrSer..... 755
    ::::| | | | | | | | | | | | | | | | | | | | | |
1026 GCAGATGGCGCAGATCGATTGCGGAAAGGAAATCCGCCCTGACGACA 1075
    ::::| | | | | | | | | | | | | | | | | | | | | |
756 .....ThrThrgluserSerSerAla 762
    ::::| | | | | | | | | | | | | | | | | | | | | |
1076 ATTTGCGGATGCGGACATACGCAATACCGTCCCTTACCATTCGCA 1125
    ::::| | | | | | | | | | | | | | | | | | | | | |
763 ProValPro.....ThrProSerSerThrThrgluserSer 776
    ::::| | | | | | | | | | | | | | | | | | | | | |
1126 AATATCCGTTCAACTGAGACGCTTACGCGCAAAAGAAACATCACCTC 1175
    ::::| | | | | | | | | | | | | | | | | | | | | |
776 lalProValProThrProSerSerThrThrgluserSerSerAla 793
    ::::| | | | | | | | | | | | | | | | | | | | | |
1176 CTCACCGGTG...CCGCGTCAAGCGAAGAAAGANTGGAACGCAACAA 1224
    ::::| | | | | | | | | | | | | | | | | | | | | |
793 roValProThrProSerSerThrThrgluserSerValAlaProVal 809
    ::::| | | | | | | | | | | | | | | | | | | | | |
1225 CGCCACCGAAGACCAAGTCCGTTGACGCTAAAGGTTCCGAATT 1274
    ::::| | | | | | | | | | | | | | | | | | | | | |
810 ProThrProSer..... 814
    ::::| | | | | | | | | | | | | | | | | | | | | |
1275 TGAAAAAGAGCTAAATACGATACGAAATTAATACCGCTGATCCGCAAG 1324
    ::::| | | | | | | | | | | | | | | | | | | | | |

```

```

815 .....SerSerAsnIleThrSerSerAlaProSerS 825
1325 TGAATCCATAGATGAA.....CCCGCTTTAAT 1353
      |||||
825 erThrProPheSerSerSerThrGluSerSerSerValProValProThr 841
1354 CCTAAGAGTTCTGTC.....GGATCGGCTCATCTTGCTATTAAC 1394
      |||||
842 ProSerSerSerThrThrGluSerSerSerAlaProValSerSerSerThr 858
1395 TGGCAGATTCATATACGAAATTCACAGGACAGTAGATCAGATATATA 1444
      |||||
858 rThrGluSerSerValAlaProValProThrProSer..... 870
1445 TCCGACCTAAATAATTACTCTCTTACGACCGCTACCAAAAGACCTAAT 1494
      |||||
871 ..SerSerSerAsnIleThrSerSerAlaPro...SerSerIleProPhe 885
1495 AATGATATTTGGATTAATTTGGTAATGATGACTTAAGGTCATCAAG 1544
      |||||
886 SerSerThrThrGluSerPheSerThrGlyThrThrValThrProSerSe 902
1545 AACTAAGGTCACAGATTTGAATGGATGTCAATGTCTAAACA 1590
      |||||
902 rSerIleTyrProGlySerGlnThrGluThrSerValSerSerThr 917

```

seq_name: p1r2:T29634

```

seq_documentation_block:
hypothetical protein C12D12.1 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 17-Mar-2000
C:Accession: T29634
R:Man, M.; Hawkins, J.
submitted to the EMBL Data Library, March 1996
A:Description: The sequence of C. elegans cosmid C12D12.
A:Reference number: Z20656
A:Accession: T29634
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1 825 <NHA>
A:Cross-references: EMBL:U51998; PID:AAA96080.1; GSPDB:GN00028; CESP:C12D12.1
A:Experimental source: strain Bristol NZ; clone C12D12
C:Genetics:
A:Gene: CESP:C12D12.1
A:Map position: X
A:Introns: 48/1: 86/3: 137/1: 172/3: 224/3: 253/1: 287/3: 328/2: 454/1: 487/3: 692/1
C:Superfamily: Epstein-Barr virus membrane antigen gp350

```

```

alignment_scores:
Quality: 146.50      Length: 401
Ratio: 0.763        Gaps: 18
Percent Similarity: 47.880      Percent Identity: 22.195

```

alignment_block:

US-09-303-518D-465 x T29634 ..

```

Align seg 1/1 to: T29634 from: 1 to: 825
281 TTTCGATCAG.....GGCAGAGTCATCCCTCGACAAACAT 324
      |||||
470 TTPProValThrThrValSerThrMetSerProThrThrThrValThrVa 486
325 GCCTCATTCGATTCGTATG.....AAGCGGTAG 356
      |||||
486 lProThrThrProThrProValProThrThrThrAsnThrProAla 503
357 TCCCGTTCGAGATGACGCTTACCGATTCATGGAGGATACGAC 406
      |||||
503 snProThrThrAlaThrProThrThrValGlyThrSerIleGlnThrAsn 519
407 ACATCCGCGG.....ACGGCTATGACGGGACAGGGCGGCTAT 450

```

```

||||| ||| .....
520 ThrIleSerProHisLeuSerThrIleThrGlySerIleValThrSerTh 536
451 CCCGCTCCCAAGGCCGACGATATATACAGCTTCGACATAAAGGCGT 500
      |||
536 rProThrMetAlaProGlnThrSerAlaSerProThr.....T 550
501 TGGCCAAATATCCGCTCAACCTGACCGACAGACCGACCGACAAAC 550
      |||
550 hProThrHisThrThrAlaSerGlnProThrThrThrLysProValVal 566
551 GCGTGTGACCGTT.....TCCAAATACCGGTAGT 582
567 ThrThrAsnSerValThrProSerThrGlyThrThrThrValProValPr 583
583 ATGTGACGACCAAGAGTAGCGACGATTCACACCGCA.....CCCG 626
583 oThrThrThrGlySerProThrThrGlnThrThrAlaProValThrLysP 600
627 ATACAGCCCGGAGCTGGACAGATCGGGCAATCCGCCGAGCTTCAACG 676
      |||||
600 roThrValProSerSerThrThrGlnThrAlaProPro...ValThrThr 615
677 GCACTCAGATATCGTCAAAATCATCATCGGCGCGACGAGAAATGTC 726
616 ProThrSerGlnProProValThrThrThrSerLeuThrThrLeuTh 632
727 GCGCAGCGGATGCGCTGCAGGATATACGGAAGGCTCAACATTCCTGT 776
632 rThrProThrValProVal.....ThrThrThrValValP 644
777 TATGACGCGCTTGGTC...TGGTTCCACCGCAAAACAGATGGCGCGCA 823
      |||||
644 roSerSerAlaThrValProThrThrProProThrThrValThrValAla 660
824 TCAACGATTTGGCAGATATGCGCACTCAAGACTATGCGCGACGCC 873
661 AlaThrThrThrSer.....LysAlaProValValThrThrSerPr 674
874 ATCCGCGATTTGGCAGATTCCAAAACCCCAATGCGGACAGCAAGCTA 923
      |||||
924 CGTCAGCATATCTTTACGCGACTCATCCCGCTCAAAAGGATGGAGCTG 973
689 roSerThrValGly.....ThrSerProThrAlaProAlaAsnLeu 702
974 TTCGGGAAATATACGCGCTTGGCGGACATCAGCAGCATCTGTCAAGCG 1023
703 ThrThrProThrThrAla..... 708
1024 TCGCAGATGGCGGATCGCATTTGCCGAAGGGAATCCGCGTCAGCA 1073
709 .....ProValAsnProThrSerSerT 716
1074 CAATTTGCGATGCGG.....CATACGCCCAATACCGCTGCC 1111
      |||
716 hThrAlaProThrAlaProValAsnProThrSerProThrThrAlaPro 732
1112 CTATACATTCGCCGAATATCCGCTTCAACTGGACGACGCTTACGGA 1161
      |||||
733 ...ThrValProProValThrThrThrThrProThrThrThrSerTh 748
1162 GAAACATCACTCTCTCAACGCTGCGCGCGCAACGAAAGATGTGAA 1211
      |||
748 rThrThr...ThrThrThrThrThrThrThrThr..... 758
1212 ACTGCAAAACAAACGACCGCAAGAACCAAAAGTCCCTTGACGGTAAG 1261
      |||||
759 .....ThrGlnThrThrProThrThrProValVal..... 768
1262 GGTTCGAATTTGAAAGAGAGTAATAATACGATACGAGATTAATAC 1311
      |||

```

769ThrThrProSerThr 773
 1312 GGTGACCAAGATGATCTATAGATGACCGCTTTATCTTAAGG 1361
 :|||:| :|||:| :|||:| :|||:| :|||:| :|||:|
 774 ILeThrProThrThrArgProValThrGlnProAlaSerGlnProAlaThr 790
 1362 TTTCTGCGATGGCTCATTTGGCTATTAAGTCCAGAAATCAATACG 1411
 :|||:| :|||:| :|||:| :|||:| :|||:| :|||:|
 790 rGlnProAlaThrThrProThrThrThrThrThrThrThrThrThr 807
 1412 CA 1413
 807 er 807

seq_name: p1r2:s57180

seq_documentation_block:

probable membrane protein YJR151c - yeast (Saccharomyces cerevisiae)
 N:Alternate names: protein J2223; serine/threonine-rich protein YJR151c
 C:Species: Saccharomyces cerevisiae
 C>Date: 23-Aug-1995 #sequence_revision 08-Sep-1995 #text_change 05-Nov-1999
 C:Accession: S57180
 R:Source: T.
 submitted to the Protein Sequence Database, September 1995
 A:Reference number: S57169
 A:Accession: S57180
 A:Molecule type: DNA
 A:Residues: 1-1161 <SCA>
 A:Cross-references: EMBL:Z49651; NID:g1015902; PID:g1015903; GSPDB:GN00010; MIPS:YJR151c
 C:Genetics:
 A:Gene: MIPS:YJR151c
 A:Map position: 10R
 C:Keywords: transmembrane protein

alignment_scores:
 Quality: 141.00 Length: 436
 Ratio: 0.610 Gaps: 16
 Percent Similarity: 52.982 Percent Identity: 22.477

alignment_block:

US-09-303-518d-465 x S57180 ..

Align seg 1/1 to: S57180 from: 1 to: 1161

71 CCTCATATTGGCAACGATCTTTATCGGAGGTTCTGACGGCTAG 120
 :|||:| :|||:| :|||:| :|||:| :|||:| :|||:|
 106 ProAlaIleSerSerAlaLeuSerIleSerIleSerIleSerIleSer 122
 121 CATTTGCAACCGGAGGAAATACCACTATTGGGAGGAGGGGAACT 170
 :|||:| :|||:| :|||:| :|||:| :|||:| :|||:|
 122 oThrSer.....ThrSerThrThrThrThrThrThrThrThrThr 137
 171 TGCCGAGGCGAGGGTCATATCGGATTTGGAAACATACAAAGCCATC 220
 :|||:| :|||:| :|||:| :|||:| :|||:| :|||:|
 137 hrProThrThrThrThrThrThrThrThrThrThrThrThrThrThr 152
 221 TGGCAACCTGTTTCATCCAGACGGGCGCATTAAGAAATATCGGCTAC 270
 :|||:| :|||:| :|||:| :|||:| :|||:| :|||:|
 153 ProThrThrThrThrThrThrThrThrThrThrThrThrThrThrThr 166
 271 ATTGTCGCTTTTCGATCAGGCGAGCAAGTCATTCCTCCCT.....T 314
 :|||:| :|||:| :|||:| :|||:| :|||:| :|||:|
 166 rThrSerThrThrProThrThrThrThrThrThrThrThrThrThrThr 183
 315 CGACAACCATGGCTCATCTCGAATTCGATGAGAGCCGTAGTCCCGT 364
 :|||:| :|||:| :|||:| :|||:| :|||:| :|||:|
 183 erThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThr 199
 365 ACGATTACGCTTACCGATCCATTTGGAGCGATGACAAACACCATCC 414
 :|||:| :|||:| :|||:| :|||:| :|||:| :|||:|
 200 ThrSerThrThrThrThrThrThrThrThrThrThrThrThrThrThrThr 216

415 GCGGAGGCTATGACGGGCGACAGGGCGGCTATCCGCTCCAAAG 464
 :|||:| :|||:| :|||:| :|||:| :|||:| :|||:|
 216 oThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThr 232
 465 GCGGAGGATATATACAGCTACGACATAAAGCGTTGCCAAATATCC 514
 :|||:| :|||:| :|||:| :|||:| :|||:| :|||:|
 233ThrSerThrThrThrThrThrThrThrThrThrThrThrThr 244
 515 GCGCAACCTGACGACACCGGACGACCGGACGACGCTTGACAGCT 564
 :|||:| :|||:| :|||:| :|||:| :|||:| :|||:|
 245 LysSerThrThrProThrThrThrThrThrThrThrThrThrThrThr 260
 565 TTCCACATACCGGTAGTATGCTGACCGAAGAGACTAGCGACGATTC 614
 :|||:| :|||:| :|||:| :|||:| :|||:| :|||:|
 260 rSerThrThrProThrThr.....SerThrThrSerThr 271
 615 ACGGCGCACCGGATATACAGCCCGGAGCTGACAGATCG..... 652
 :|||:| :|||:| :|||:| :|||:| :|||:| :|||:|
 271 hrAlaProThrThrThrThrThrThrThrThrThrThrThrThrThr 287
 653 GCATGGCGCGGAGCTTTCAAGCGGCTGACATATATGCTCAAAACATC 702
 :|||:| :|||:| :|||:| :|||:| :|||:| :|||:|
 288 SerThrAlaProThrThrThrThrThrThrThrThrThrThrThrThr 304
 703 ATCGGCGCGGAGGAGAAATTTGCGGCGGAGCGATCGCTGACAGGTAT 752
 :|||:| :|||:| :|||:| :|||:| :|||:| :|||:|
 304 rAlaSerAlaSerSerValIleSerThrThrAlaThrThrThrThrThr 321
 753 AAGGGAAGGCTCAACATTTGCTGTATGACAGGCTGGGCTGCTCTTCA 802
 :|||:| :|||:| :|||:| :|||:| :|||:| :|||:|
 321 heAlaSerLeuThrThrProAlaThrSerThrAla.....SerThr 334
 803 CCGAAACAGATGCGCGCATCAGCATTTGGCAGATATGCGGCAATC 852
 :|||:| :|||:| :|||:| :|||:| :|||:| :|||:|
 335 AspHisThrThrSerSerValSerThrThrAsnAlaThrThrThrThr 351
 853 AAAGACTATGCGCGACGACCCGATTTGGGACGACATCCAAACCCAA 902
 :|||:| :|||:| :|||:| :|||:| :|||:| :|||:|
 351 aThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThr 368
 903 TGCCGCAAGGATAGAGCCGCTGACGACAAATTTTGGCGAGTATCC 952
 :|||:| :|||:| :|||:| :|||:| :|||:| :|||:|
 368 InValThrSerSerAlaGluProThrThrValSerGluValThrSer 384
 953CCGCAAAAGGATTTGAGCTGTTGGGGAATATCGGCTGGCG 996
 :|||:| :|||:| :|||:| :|||:| :|||:| :|||:|
 385 ValGluProThrThrArgSerSerGluVal.....Thr 394
 997 GGCATCAGGCGACATCTGTCAAGCGGTGCGAGATGGCGAGATCGCAT 1046
 :|||:| :|||:| :|||:| :|||:| :|||:| :|||:|
 394 rSerSer..AlaGluProThrThrValSerGluPheThrSerSerValG 410
 1047 GCGCAAAAGGAAATCGCGCTGACGCGCAATTTTGGCGATGCGGCTAT 1096
 :|||:| :|||:| :|||:| :|||:| :|||:| :|||:|
 410 uproThrArgSerSerGluValThrSerSerAlaGluProThrThrVal 427
 1097 CCAATATCCCGTCC.....CCTTACCATTCGCGAATATCCGTTCA 1137
 :|||:| :|||:| :|||:| :|||:| :|||:| :|||:|
 427 erGluPheThrThrSerValGluProThrThrArgSerSerGluValThr 443
 1138 AACTTGAGCAGCGTTACGCAAGAAAGAAACATCACCCTCTCAACCGTGC 1187
 :|||:| :|||:| :|||:| :|||:| :|||:| :|||:|
 444 SerAlaGluProThrThrValSerGlu...PheThrSerSerValGlu 459
 1188 GCGCTCAAGGAAAGATGTGAAGTGGCAACAAACCGCCACCGCAAGA 1237
 :|||:| :|||:| :|||:| :|||:| :|||:| :|||:|
 459 oThr.....ArgSerSerGluValThrSerSerAlaGluProThrThr 473
 1238 CCAAGTGCCTTTGAGCGTAAAGGCTTCCGAAATTTGAAAGAGCGTA 1287
 :|||:| :|||:| :|||:| :|||:| :|||:| :|||:|
 473 hrValSerGluPheThrThrSerSerValGluProThrThrArgSerSer 489
 1288 AAATACGAT.....ACGAGAAATTAATACCGCTGTACCAACAACTGAA 1328


```

454  InProValProGlyThrLysArgThrLysSerAsnTyrcysGlu..... 468
1153  TACGGCAAGAAACATCACCTCTCTCAACCGTCCGCCGCTCAACAGGAAA 1202
469  ....AsnGluProAsnLysSerSerGlnSerLeuValSerSerGluSerH 484
1203  GAATGTGAACCTGGCAACCAAGCCGACCCGAGACCAAGATCGCGTTTG 1252
484  sAsnValGluGlyTrpAsnTyrcysGluThr..... 496
1253  ACGGTAAAGCGTTTCCGAATTTGAAAAAGAGCTAAATACGAT..... 1296
497  ....ValGlyPhe.....TyrAspProSer 503
1297  ACGAATATATACCCGCTACCAAGTATGATTCATATGATCAACCCCT 1346
504  AlagLuleSerAlaSerLeuAspGluLeuArgGln...SerThrProVa 519
1347  CTTTAATCTTAAGGTTCTGTCGATCGGCTCATCTTCTGCTATTAATG 1396
519  lAlaArgAspSerGluLeuLeuSerArgAlaHisSerPheAspLeuAsn 535
1397  CCAGAATTCAATACGCAAAATTCACAGCAGGATAGATTCAGATATATC 1446
536  ..ArgLeuAsp.....Leu 539
1447  CCACCTTAATAATTAATCTCTCCCTACGACCGCTACCAAAAGACATATA 1496
540  ProSerGlnAspLysSerThrSerTyrgluValProAsnGlyThrGluAs 556
1497  T.....CGATATTTGGATAAA...TTTG 1516
556  ngInSerProArgProValThrSerLeuGlyPheValAsnGluThrPheP 573
1517  GTAATGAATGACTAAGCTCA 1539
573  heGluGluLysProLysAlaPro 580

seq_name: p1r2:JT0345

seq_documentation_block:
  description: (EC 2.4.1.5) precursor - Streptococcus mutans (strain GS-5)
  N:Alternate names: sucrose 6-glucosyltransferase
  C:Species: Streptococcus mutans
  C:Date: 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change 24-Sep-1999
  C:Accession: JT0345; C33135
  R:Ueda, S.; Shiroza, T.; Kuramitsu, H.K.
  Gene 69, 101-109, 1988
  A:Title: Sequence analysis of the gtfC gene from Streptococcus mutans GS-5.
  A:Reference number: JT0345; MUID:89137980
  A:Accession: JT0345
  A:Molecule type: DNA
  A:Residues: 1-1375 <UED>
  R:Shiroza, T.; Ueda, S.; Kuramitsu, H.K.
  J. Bacteriol. 169, 4263-4270, 1987
  A:Title: Sequence analysis of the gtfB gene from Streptococcus mutans.
  A:Reference number: A33135; MUID:87308013
  A:Accession: C33135
  A>Status: Preliminary
  A:Molecule type: DNA
  A:Residues: 1-349 <SHI>
  A:Cross-references: GB:M17361
  C:Genetics:
  A:Gene: gtfC
  C:Function:
  A:Description: catalyzes the synthesis of both water-soluble and water-insoluble glucans
  C:Superfamily: cpl repeat homology
  C:Keywords: duplication; glycosyltransferase; hexosyltransferase
  F:1-34/Domain: signal sequence #status predicted <SIG>
  F:35-1375/Product: glucosyltransferase #status predicted <ANT>
  F:1126-1145/Domain: cpl repeat homology <CP1>

```

```

F:1253-1272/Domain: cpl repeat homology <CP2>
F:1318-1337/Domain: cpl repeat homology <CP3>

alignment_scores:
  Quality: 140.00      Length: 606
  Ratio: 0.507        Gaps: 32
  Percent Similarity: 45.545      Percent Identity: 22.112

alignment_block:
  US-09-303-518D-465 x JT0345

Align seg 1/1 to: JT0345 from: 1 to: 1375

64  GCACACGCTCAGATTGCGCAAGATTTGATTCGTCGACGAGTTCTGCA 113
    :::::||||| :::::||||| :::::|||||
643  SerTyralaLeuLeuLeuThrAsnLysSerSerValProArgValTyty 659
    :::::||||| :::::||||| :::::|||||
114  CCGTCAGCATTTGCAACCCGACGGCAATAC..... 144
    :::::||||| :::::|||||
659  rGlyAspMetPheThrAspAspGlyIntTyMetAlaHisLysThrIleA 676
    :::::||||| :::::|||||
145  .....CACCTATTC 153
676  snTyrgluAlaIleGluThrLeuLeuLysAlaArgIleLysTyValSer 692
    :::::||||| :::::|||||
154  GGCAGCAGGGGGGAACTTCCGACGCGGCTCAT..... 189
    :::::||||| :::::|||||
693  GlyGlyGlnAlaMetArgAsnGlnGlnValGlyAsnSerGluIleLeTh 709
    :::::||||| :::::|||||
190  .....ATCGATTGGGAACATCAAAAGCATCAGTTGGGCAAC 229
    ||| ||| :::::|||||
709  rSerValaTyrglyGlyLysGlyAlaLeuLysAlaThrAspThrGlyAspA 726
    :::::||||| :::::|||||
230  TGTTCAATCCAGCAG.....GGCGCATTAAGAAATATCGGCTAC 270
    :::::||||| :::::|||||
726  rGThrThrArgThrSerGlyValAlaValIleGluGlyAsnAspProSer 742
    :::::||||| :::::|||||
271  ATP.....GTCCGCTTTCCGAT.....CA 290
    :::::||||| :::::|||||
743  LeuArgLeuLysAlaSerAspArgValAlaValAsnMetGlyAlaAlaH 759
    :::::||||| :::::|||||
291  CGGCGACGAAATCCATTCGCCCTTC.....GACACAC.....C 322
    :::::||||| :::::|||||
759  sLysAsnGlnAlaTyArgProLeuLeuLeuThrThrAspAsnLysIleL 776
    :::::||||| :::::|||||
323  ATGCCTCAATTCGATCTCTGATGAAGCCGATGCTCCGTTGACGATTC 372
    ||| ||| :::::|||||
776  yAlaTyThrIleSerAspGlnGluAlaIaGly..... 786
    :::::||||| :::::|||||
373  AGCCTTTACCGCATTCATTGGAGCGATACGA.....CACCATCCGC 416
    ||| ||| :::::|||||
787  ..LeuValArgTyThrAsnAspArgGlyGluLeuIlePheThrAlaAl 802
    :::::||||| :::::|||||
417  CGAC.....GGCTATGACGGGCGACAGGGGGGCTATCC..... 453
    ||| ||| :::::|||||
802  AspIleLysGlyTyAlaAsnProGlnValSerGlyTyLeuGlyValT 819
    :::::||||| :::::|||||
454  ..GCTCCCAAGGGCGGAGGATATATACAGCTACGACATAAAGCGCTT 501
    :::::||||| :::::|||||
819  rValaProValGlyAla.....AlaAla 826
    :::::||||| :::::|||||
502  GCCCAAAATATCCGCTCAACCTGACCGACAGACCGACCGGACAGACG 551
    :::::||||| :::::|||||
827  AspGlnAspValaArgValAlaAlaSerThrAlaProSerThrAspGly 843
    :::::||||| :::::|||||
552  GCTGTGACCGTTTCCACATATACCGGTAGTATGCTGACCAAGAGATAG 601
    :::::||||| :::::|||||
843  sSerVal.....HisGlnAsnAlaAlaLeuAspSerArgValaMetP 857
    :::::||||| :::::|||||
602  GCGACGATTCMAAGCGCACCGATACAGCCCGGCGGTGACAGATAG 651
    :::::||||| :::::|||||
857  heGluGlyPheSerAsnPheGlnAlaPheAlaThrLysLysGluGluTy 873
    :::::||||| :::::|||||

```

```

652 GGCATGCCCGCGAAGCTTTCAC.....GGCAC 680
    |||:||||| ||| |||
874 ThrAsnValValIleAlaLysAsnValAspLysPheAlaGluProGlyIle 890
    |||:||||| ||| |||
681 TGCAGATATGTCATAAAACATCATCGCGCGGAGGAATATGTGGCG 730
    |||:||||| ||| |||
890 ThrAsp.....PheGluMetAlaProGluTyrValSer 902
    |||:||||| ||| |||
731 CAGGCTATTAAGCTCAAGCAAGCTCAACAT 771
    |||:||||| ||| |||
902 ThrThrAspGlySerPheLeuAspSerValIleGluAsnGlyTyrAlaPhe 918
    |||:||||| ||| |||
772 GCTGTATGCAAGCTTGGCTGTGCTTTCACCGAACAAGATGGCGC 821
    |||:||||| ||| |||
919 ThrAspArgTyrAsnLeuGly...IleSerLysProAsnLysTyrGly 934
    |||:||||| ||| |||
822 CATCAAGATTTGGCAGAT...ATGGCGCAACTCAAGACTATGCGCGC 868
    |||:||||| ||| |||
934 rAlaAspAsnLeuValLysAlaIleLysAlaLeuHisSerLysGlyIle 951
    |||:||||| ||| |||
869 CAGCATTCGCGATTTGGGAGTCACAAACCCCAATGCCGCAAGCATA 918
    |||:||||| ||| |||
951 yValMetAlaAspTyrValProAspGluMetTyrAlaLeuProGluLys 967
    |||:||||| ||| |||
919 GAAGCCGTACAGCAATCTTTACGCGACTCATCCCGTCAAGGATGG 968
    |||:||||| ||| |||
968 GluValValThr..... 971
    |||:||||| ||| |||
969 AGCTGTTCGG...GGAAATACGGCTTGGGCGCATCAGCGACATCTG 1015
    |||:||||| ||| |||
972 .AlaThrTyrValAspLysTyrGlyThr.....ProV 982
    |||:||||| ||| |||
1016 TCACGCGTGCAGATG...GGCAGATCGCATTCGCAAGGAATCC 1062
    |||:||||| ||| |||
982 AlaIleGlySerGluIleLysAsnThrLeuTyrValValAspGlyLysSer 998
    |||:||||| ||| |||
1063 GCCGTCACCGCAAT.....TTTCCGCAATGCGGC 1091
    |||:||||| ||| |||
999 SerGlyLysAspGluGlnAlaLysTyrGlyValAlaPheLeuGluGlu 1015
    |||:||||| ||| |||
1015 uGlnAlaLysTyrProGluLeuPheAlaArgGluIle..... 1028
    |||:||||| ||| |||
1142 TGGACAGCGGTACGGCAAAACATCACTCTCTCAACCGTGGCGCG 1191
    |||:||||| ||| |||
1029 .....SerThrGlyValProMet 1034
    |||:||||| ||| |||
1192 TCACACGGAAGATGTGAA...CTGGCAACAAACCCACCCGAGAC 1238
    |||:||||| ||| |||
1035 AspProSerValLysIleLysGluIlePheSerAlaLysTyrPheAsnGly 1051
    |||:||||| ||| |||
1239 CAAAGTCCGTTGACGGTAAAGGTTTCCGCAATTTGAAAAAGCCTTA 1288
    |||:||||| ||| |||
1051 rAsnIleLeuGlyArgGlyAlaGlyTyr.....ValL 1062
    |||:||||| ||| |||
1289 AATAGATACGAGAAATTATACCGCTGACCAAGATGATCTATAT 1338
    |||:||||| ||| |||
1062 euLysAspGlnAlaThrAsnThr.....TyrPheSerLeuValSer 1075
    |||:||||| ||| |||
1339 GAACCGCTTTAATCTTAAGGTTCTGCGATCGGCTCAT.....TC 1382
    |||:||||| ||| |||
1076 AspAsnThrPheLeuProLysSerLeuValAsnProAsnHisGlyThrSe 1092
    |||:||||| ||| |||
1383 TTGCTTAATACCTCCAGAAATTCATACCAAGCAAGCAAGTA 1432
    |||:||||| ||| |||
1092 rSerSerValThrGlyLeuValPheAspGlyLys.....GlyT 1105
    |||:||||| ||| |||
1433 GAATCAGATATATCCACCTAAATATCTCTCTCAGCAGCGCTACCA 1482
    |||:||||| ||| |||
1105 yValLysTyrSerThrSerGlyAsnGlnAlaLysAsnAlaPheIleSer 1121
    |||:||||| ||| |||

```

```

1483 AAGGACCTAATATGATATTTCGATAAATTTGTATGATGACTAA 1532
    ||| ||| |||:||||| ||| |||
1122 LeuGlyAsnAsnTyrPyrThrPheAspAsnAsnGlyTyrMetValThr.. 1137
    |||:||||| ||| |||
1533 AGCTCCATCAAGAACTAAAGCTCAAGAAATTTGAATGG.....GATG 1573
    |||:||||| ||| |||
1138 .GlyAlaGlnSerIleAsnGlyAlaAsnTyrThrPheLeuSerAsnGlyI 1154
    |||:||||| ||| |||
1574 TTCATTTGTCTAAA.....ACAGAGACAGACACTTGATGG 1611
    |||:||||| ||| |||
1154 LeuIleLeuArgAsnAlaIleTyrAspAsnGlyAsnLysValLeuSerTyr 1170
    |||:||||| ||| |||
1612 GCTAGTAGGATGTATAG 1629
    |||:||||| ||| |||
1171 TyrGlyAsnAspGlyArg 1176
    |||:||||| ||| |||

seq_name: pir2:J00405

seq_documentation_block:
hypothetical 119.5K protein (uvra region) - Micrococcus luteus
N:Alternate names: ORF 1 protein
C:Species: Micrococcus luteus, Micrococcus lysodeikticus
C:Date: 31-Dec-1990 #sequence_revision 31-Dec-1990 #text_change 04-Feb-2000
C:Accession: J00405
R:Shiota, S.; Nakayama, H.
Mol. Gen. Genet. 217, 332-340, 1989
A:Title: Micrococcus luteus homolog of the Escherichia coli uvra gene: identification
A:Reference number: S04781; MUID:89364717
A:Accession: J00405
A:Molecule type: DNA
A:Residues: 1-1106 <SH1>
A:Cross-references: EMBL:X15867
A>Note: this reading frame extends between two stop codons and does not begin with a
C:Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homolo

alignment_scores:
Quality: 139.50 Length: 638
Ratio: 0.524 Gaps: 37
Percent Similarity: 41.693 Percent Identity: 24.295

alignment_block:
US-09-303-518D-465 x J00405 ..

Align seg 1/1 to: J00405 from: 1 to: 1106

30 TGTGTCATACGTGCAAGTGTGCTGCGCATGTCACACCGCTCAGATT 79
    |||:||||| ||| |||
116 SerValArgLeuGlnValLeuProGlyLeuArgHisAspLeuArgAr 132
    |||:||||| ||| |||
80 TGGCAACGATCTTTTATCCGCGCAGGTTCTGCACCGTCAAGCTTCGAA 129
    |||:||||| ||| |||
132 gGly.....ProAlaIleLeuArgValAlaLeuLeu 144
    |||:||||| ||| |||
130 CCGGACGGAAATACACCTATTCGGCAGCGAGGGGCAACTTCCGAGCG 179
    |||:||||| ||| |||
144 euArgProHisValProGlyLysGlnIleValArg...GlyLeu 159
    |||:||||| ||| |||
180 CAGCGGTATATCGGATTTGGGAACATACAAAGCATCACTATGGCAACC 229
    |||:||||| ||| |||
160 HisArgGlyProValProGlyArgValHisArgProGluValHisGlnP 176
    |||:||||| ||| |||
230 TGTTCATCCGCA.....GGCGGCATTAAGGAATATACGCGCTAC 270
    |||:||||| ||| |||
176 o...GlnProAlaLeuHisSerGlyAspHisHisArgAspLeuArgLeu 192
    |||:||||| ||| |||
271 ATGTCCGCTTTCCGATCA..... 290
    |||:||||| ||| |||
192 isAlaProAlaLeuGlyThrCysArgGlyAlaAlaLeuProAlaValArg 208
    |||:||||| ||| |||
291 CGGCGACGAACTCATTC.....CCCTTCGACA 319
    |||:||||| ||| |||

```

```

209 ArgAlaGlyGluProAlaAspProAlaAlaAspArgGlyProAlaArgGly 225
320 ACCATGCTCCATTCGATCTGTGATGAAGCGGTAGTCCGTTGACGGA 369
    ::::
225 yAlaAla.....Ala 229
370 TTCAGCGCTTA.....CCGATCCATTGGAGCG 398
    ::::
229 LAlHsProLeuProGlyAlaArgGlyProArgProGlnGlyArg 245
399 ATACGAAACCATCCCGCGAGCGGTATGACGGCCACAGGCGCGCGCT 448
    ::::
246 ValArgGlyProValGlnGlyProVal...HsAlaGlyLeuAlaArgArg 261
449 ATCCGCGCTCCAAAGG.....CCGAGG 471
    ::::
261 gGlyArgGlyAspArgProAlaLeuGlyProAlaArgAlaGlnGlyAlaG 278
472 GATATATACAGCTACGACATAAAAGCGCTGCCAAA..... 509
    ::::
278 LylAlaAlaHsHsArgArgArgGlyProProArgHsGlnGlyGly 294
510 TATCGCGCTCAACCTGACCGACACACCGGACCGGACAGCGCTTTCG 559
    ::::
295 HsProProAlaPro.....HsArgLeuGlyGlyAspArg 306
560 ACCGTTTCCACATACCGGTAGTAGTGCACGACGAGTAGCGACGGA 609
    ::::
306 gPro.....GlnAlaGlyGlyArg..... 312
610 TTCMAACGCCACCCGATACGCCCGGACGTGACAGATC..... 650
    ::::
313 .....ProArgArgGlyArgValArgGlyArg 321
651 GGGCAATGCCCGGAGCTTTCACAGCGACCTGCAGATATTCGCAAAACA 700
    ::::
322 GylAlaGlyArgArgGlnGlyGlnGlnHsArgGlyValArgArgAl 338
701 TCATCGCGCGCGC.....AGGAAATTTGTC 726
    ::::
338 aGlyArgGlyGlnProProValProLeuValLeuArgGlnAlaLeuL 355
727 GCGGACGAGGATCGCGTGCAGGGTATTAAGCCAGGCTCAACATTGCTGT 776
    ::::
355 euProGlnArgAlaArgAla..... 361
777 TATGCAGCGCTGGGTCTGCTTCCACGAAACAAGATGGCG..... 821
    ::::
362 .....AspArgGlyArgAspArgAlaAlaLe 370
822 .....CATCAAGATTTGGCAGATATGGCGCAACTCAAAAGACTAT 861
    ::::
370 uValLeuGlnGlnProValArgArg.....ValProArgValH 384
862 GCGGACGACGATCCGGATGGGACATCCAAACCCCAATGCCGACA 911
    ::::
384 LsArgHsArgLeuProProAlaGlyGlyProGlyProArgArgArgGln 400
912 AGGCAT.....AGAACCGCTACGAAATCTTTACGGGACGCA 949
    ::::
401 ArgArgAlaValAlaProAlaArgGlyArgArgArgAlaValAlaAlaArg 417
950 TCCCGGTCAAAGGATGGAGCTGTTCCGGGAAAA.....TACGGC 990
    ::::
417 nValHsLsLeuGlyLeuLeuAlaAlaArgAlaArgAlaGlyGlnGly 433
991 TTGGGCGCATCAGCGCACATCTCTCAAGCGGT..... 1024
    ::::
434 AspGlyLeuLeuProGlnHsProValGlnGlyProAspGlyGlyGly 450
1025 .....CGAGATGGCGAGATCGCATCCGAAAGCGCAAT..... 1060
    ::::
450 LArgArgArgProAlaArgGlnGlyLeuGlnGlyGlyAspValPro 466

```

```

1061 CCGCGCTCAGGACATTTTCCGATCGGCGATACGCCAATACCGGT... 1108
    ::::
467 GlnProValArgProAlaArgAlaProLeuHsHsGlyLeuArgGlyArgH 483
1109 .CCCGTTACCATTTCCGAAATATCC..... 1132
    ::::
483 sProLeuArgAspAlaGlnAlaArgGlyAspArgValGlyArgArgPro 500
1133 .....GTCAAACTTGGACGACGCTTACGGCAAAAGAAACA 1168
    ::::
500 rGAlaLeuArgValAlaHsAlaGlyAspProValProGlyValProArg 516
1169 TCACCTCTCCACCGCTG.CCGCGCTCAAGCGAAAGAAATGTAACCGCG 1217
    ::::
517 GylProProGlnProHsGlyProGlnArgAlaArgGlyArgProValH 533
1218 AATCAACGCCACCCGAAAGACAAA..... 1242
    ::::
533 sArgGlyArgHsProProAlaHsAlaArgGlyHsGlyValLeuLeuG 550
1243 .TGCGCGTTTGACGCTAAAGGCTT.....CCGAATTTTGA 1278
    ::::
550 LylAlaAlaAlaAspGlyProGlyAlaAlaAspArgGlyProGlyAlaGln 566
1279 AAAGAC.....GTAAATACGATACGAGATTAATACCGC 1313
    ::::
567 GylAspProGlyProAlaGlyValProAlaGlyArgArg.....ProArg 581
1314 TGATCAACAAGTGAATCTATAGATGAACCGCTTTAATCTCAAAAGTT 1363
    ::::
581 gValProGlnProArgAlaAlaGlyArg.....HsProLeuArg 595
1364 CTGTCGATGGCGCTCATTTGTCATATGCTGCACGAAATTCATACGCA 1413
    ::::
595 rGArgGlyProAlaHsProProGlyHsHsArgArgLeuArgAlaGly 611
1414 AAATTCACCAAGGCAAGGTAGATCAGATATATC..... 1446
    ::::
612 ArgArgProLeuArgProArg...ArgAlaValHsArgProAlaProAl 627
1447 .....CCACSTAAATTAATCTCTCTTCAACGCGCTACCAAAAGAC 1489
    ::::
627 aGlyGlnProProHsArgAspPro.....ProAlaProAlaGly 642
1490 CT.....AATTAAT 1497
    ::::
642 rOArgGlnHsProHsArgArgAlaArgAlaArgGlyHsAspArgArg 658
1498 GGATATTTGGATAAATTTGTAATGAATGCACTAAAGCTTCATCAAGAC 1547
    ::::
659 GylGlyLeuAspArg...GlyHsArgProSerArgGlyArgValArgArg 674
1548 TAAAGTCAA 1557
    ::::
674 gArgGlyArg 677
seq_name: p1r2:A48018
seq_documentation_block:
mucin 7 precursor, salivary - human
N:Alternate names: mucin, MG2; mucin, MG2a-T1; mucin, MG2b-T2
C:Species: Homo sapiens (man)
C:Date: 16-Feb-1994 #sequence_revision 18-Nov-1994 #text_change 07-May-1999
C:Accession: A48018; S29115; S29116; S29114
R:Bobek, L.A.; Tsai, H.; Biesbrock, A.R.; Levine, M.J.
J. Biol. Chem. 268, 20563-20569, 1993
A:Title: Molecular cloning, sequence, and specificity of expression of the gene encod
A:Reference number: A48018; MUID:9338636
A:Accession: A48018
A:Molecule type: mRNA
A:Residues: 1-377 <BOB>
A:Cross-references: GB:L13283

```

A: Experimental source: submandibular gland
 A: Note: sequence extracted from NCBI backbone (NCBI:137719, NCBI:137720)
 R: Reddy, M.S.; Bobek, L.A.; Haraszthy, G.G.; Biesbrock, A.R.; Levine, M.J.
 Biochem. J. 287, 639-643, 1992
 A: Title: Structural features of the low-molecular-mass human salivary mucin.
 A: Reference number: S29114; MUID:93075006
 A: Accession: S29115
 A: Molecule type: mRNA
 A: Residues: 143-168 <RED>
 A: Accession: S29116
 A: Molecule type: protein
 A: Residues: 'S', '71-79', 'N', '81-86', 'XX', '89', 'X', '91', 'P' <RE2>
 A: Accession: S29114
 A: Molecule type: protein
 A: Residues: 143-145, 'X', '147', 'XXX', '151-152', 'X', '154-158', 'X', '160-161', 'A', '163-164', 'XX', '167-168'
 C: Genes: GDB: MUC7
 A: Cross-references: GDB: 138799; OMIM: 158375
 A: Map position: 4q13-4q21
 C: Keywords: glycoprotein
 F: 1-18/Domain: signal sequence #status predicted <SIG>
 F: 19-377/Product: mucin 7, salivary #status predicted <MAT>
 F: 97,128,135,146,312/Binding site: carbohydrate (asn) (covaleqt) #status predicted

alignment_scores:
 Quality: 139.00 Length: 370
 Ratio: 0.952 Gaps: 19
 Percent Similarity: 39.459 Percent Identity: 24.324

alignment_block:
 US-09-303-518D-465 x A48018 ..

Align seg 1/1 to: A48018 from: 1 to: 377

```

119 AGCATTTCGACCCGAGGGGAAATACCACTATTCGGCAGCAGGGGGGAA 168
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
99  SerValValAsnProThrLeuValAlaThrGlnIleProSerValTh 115
169 CTTGGCCGAGCGGCGTCAATGCGATGGGAAACATCAAAAGCATCA 218
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
115 rPheProSerAlaSer.....ThrLysIleThrLeuProAsnValT 130
219 GTTGG.....GCACCTGTCATCCAGCAGCGGCCATTAAAG 256
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
130 hPheLeuProGlnAsnAlaThrThrIleSerSerArgLysValAsn 146
257 GAATATCGGCTACATGTCGGCTTTCGATCAGCGGCGCAAGTCCAT 306
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
147 ThrSerSerValAlaThrLeuAlaProVal.....AsnSerProAl 161
307 TCCCCCTTCGACAAAC...ATGCTTCACATTCGATTCGATGAAGCGG 353
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
161 aProGlnAspThrThrAlaAlaProProThrProSerAlaThrProA 178
354 TAGTCCCGTTGACGAGTTCAGCCTTACCGCATTCATTGGAGCATACG 403
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
178 laProPro.....SerSerSerAlaProProGluThr 188
404 AACACATCCCGCGCGCGCTAGTACGCGGCCACAGGCGCGGCTATCC 453
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
189 ThrAlaAlaProProThrPro..... 195
454 GCTCCCAAGGCGCGAGGATATATACAGCTACGACATAAAGCGCTTGC 503
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
196 .....SerAlaThrThr.....GlnAlaP 202
504 CCAAAATATCCGCTTCACCTACGACGACGACGACGACGACGACG 553
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
202 rProSerSerSerAlaProProGluThrThrAlaAlaProProThrPro 218
554 TTTCGACCGTTCCGACATACCGGTAGTATGCTGACGCAAGAGATAGC 603
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

```

219 ProAlaThrThrProAlaProPro..... 226
604 GACGATTCGAAGCGCGCCACCGCATACA.....GCCCGAGCTGGACAG 647
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
227 .....SerSerSerAlaProProGluThrThrAlaAlaProProThrPro 242
648 ATCGGCAATGCGCGCGAGCTTTCACAGGCGACTGAGATATGTCGAAAA 697
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
242 eAlaThrThrProAlaProLeuSer..... 250
698 ACATCATCGGCGCGCGAGGAAATGTCGCGCGAGCGATGCGTCGACG 747
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
251 .....SerSerAlaProProGluThrThrAlaAlaProProThrPro..... 264
748 GGTATAGCCAGGCTTCAAACATTCGCTATGACAGCGCTTGGTCTGCT 797
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
265 .....SerAlaThrThrLeuAspProSerSerAla.....SerA 276
798 TTCACCGCAAAACAAAGATGCGCGCATCAACGATTTGGCAGATATGCGC 847
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
276 laProProGluThr..... 280
848 AACTCAAGACTATGCGCGAGCAGCAGCATCCGCGATTGGGATGCAAAAC 897
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
281 .....ThrAlaAlaProProThrProSerAla.....ThrTh 291
898 CCCAATGCGCGCAAGGCAATGAGAACGCGTCAGCAATATCTTTACGGCAGT 947
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
291 rProAlaPro.....ProS 296
948 CATCCCGCTCAAAAGGATGAGAGCTGTCGGGAAATACGCGCTTGCGCG 997
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
296 eSerProAla..... 299
998 GCATCAGGCGCATCTGTCAACGCGTCGCAATGGCGCGATCGCATTTG 1047
299 ..... 299
1048 CCGAAAGGAAATCCGCGCTCAGCAGCAATTTCCGATGCGGCAATACG 1097
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
300 .....ProGlnIleThrThrAlaAlaProIleThr.....ThrP 311
1098 CAATATCCCGCTCCCTTACCATTT...CCGCAATATCCGCTT..... 1135
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
311 roAsnSerSerProThrThrLeuAlaProAspThrSerGluThrSerAla 327
1136 .....CAACTGGAGCAGCGCTTACGCGCAAAAGAAACATCACC 1173
328 AlaProThrHisGlnThrThrThrSerValThrThrGlnThrThrThr 344
1174 TCCTCAACCG 1183
344 rLysGlnPro 347

```

seq_name: p1r2:T02345

seq_documentation_block:
 hypothetical protein KIA0324 - human (fragment)
 C: Species: Homo sapiens (man)
 C: Date: 05-Mar-1999 #sequence_revision 05-Mar-1999 #text_change 05-Nov-1999
 C: Accession: T02345
 R: Rieke, D.O.; Bruce, D.; Mundt, M.; Doggett, N.; Munk, C.; Saunders, E.; Robinson, D.
 re, J.; White, S.; Ueng, S.; Tatum, O.; Campbell, C.; Pawcett, J.; Deaven, L.
 submitted to the EMBL Data Library, March 1998
 A: Description: Sequencing of human chromosome 16p13.3.
 A: Reference number: Z14664
 A: Accession: T02345
 A: Status: preliminary; translated from GB/EMBL/DBJ
 A: Molecule type: DNA
 A: Residues: 1-1791 <RIC>
 A: Cross-references: EMBL:AC004493; NID:g2996648; PIDN:AAC08453.1; PID:g2996650
 C: Genes:
 A: Map position: 16

alignment_scores:

Quality: 137.50 Length: 572
 Ratio: 0.476 Gaps: 32
 Percent Similarity: 50.524 Percent Identity: 21.503

alignment_block:
 US-09-303-518D-465 x T39758 ..

Align seg 1/1 to: T39758 from: 1 to: 615

```

43 GCACGTGCTCCGCGATGATGCACACGCTCAGATTGGCAACGATTC 92
   |||||
   : : : : :
16 AlaValProLysSerLeuAlaGlySerThrSerAsnIleProAsnGlu 32
   : : : : :
93 TTTATCCGGCAGGTCTCGACGCTCAGCATTTGCAACCCGACGAGAAAT 142
   |||||
   : : : : :
32 upheVal.....GlyProAspLysArg 40

143 ACCACCTATTGGCAGCAGGGGGAGACTTGGCCGACGCGATATC 192
   : : : : :
40 rOleuSerGlnAsnAlaGlnIleLysSerLysSerSerLysValVal 56

193 GGATTGGGAACATACAAGCCATCAGTTGGGCAACCTGTTCAATCAGCA 242
   : : : : :
57 ProGlnAlaLeuValProGlnAspAspIle..... 66

243 GCGCGCCATTAAAGGAAATATCGGCTAC.....ATTGTCGCT 280
   |||||
   : : : : :
67 ArgAlaValAlaGlnGlyIleLeuAsnTyrGlyGlySerAsnAspAsnArg 83

281 TTTCGATCAGCGGCAC.....GAATCCATTCCTCCCTTGACAAAC 321
   : : : : :
83 rOValSerHisThrHisThrPheValGlnLeuGlnLysSerHisGlnAsn 99

322 CATGCTCAGATTCGATTCCT.....GATGAAGC 350
   |||
   : : : : :
100 HisLeuThrGlnAsnAspArgAsnPheGlyThrSerArgLeuAspAspVal 116

351 CGGTAGTCCGCTTGACGATTCAGCTTACCGCATTCATGGACGAT 400
   : : : : :
116 AlaProAsnAlaAspGlyValArgArgLeuArg..ThrSerGlySer 132

401 ACG.....AACACCATCCCGCGCGAGCTATGACGGCCACAGGGCGC 444
   |||
   : : : : :
133 ThrGlyLeuSerAsnAlaProProSerAlaAsnValSerLysAlaSer 149

445 GGCTATCCCGCTCCCAAGGCGGAGGATATATACAGCTACAGACATAA 494
   : : : : :
149 rAsnLeuSerLeu.....AlaSerLeuAlaLysThrGlnProGlnVal 163

495 AGCGCTGCGCCAAATATCCGCTCAGCTGACGACAAACGACGACA... 541
   |||||
   : : : : :
163 rGAlaThrProGlnValCysValProLeuAsnProAspThrGlySerVal 179

542 .....CCGACAAAGGCTTGTG..... 558
   |||||
180 ProLeuIleHisProGlnGlnThrAspArgGlyLeuProTyrAlaProAs 196

559 .GACGCTTTCCACATACCGGTAGTATG...CTGACGCAAGAGAGAGGC. 603
   : : : : :
196 rGlnLysPheHisAsnSerGlySerLeuLysLeuProLysGlyAlaSerL 213

604 ..GACGATTCAAACGCGCCACCCGATACAGCCCGAG.....CTG 642
   : : : : :
213 euGlnAspLeuSerArg.....SerProSerArgAlaValLeu 225

643 GACGATCGGCGCATGCGCCGAGCT..... 669
   : : : : :
226 AsnGlnAspLysAsnValAspGlnCysAlaProProGlnProTyrProL 242

670 ....TTCAAGCGCATCGCATATCTGTCAAAAC...ATCATGCGC.... 708
   : : : : :

```

```

242 nGlnTyrAsnGlnValLeuAspAspValGlnAsnAlaValValGlyThrS 259
   : : : : :
709 .....GCGGACGAGGAATGTGCGC 729

259 erProLeuGlnTyrThrSerLysProLeuAlaAlaAsnValGlnArgSer 275
   : : : : :
730 GCGAGCGATGCCGTCGACGAGTATAAGCAGGAGCTCAACATTCGCTGTAAT 779
   : : : : :
276 ThrAlaAsp.....LeuThrGlnSerAspAsnIleCys..... 286

780 GCACGCTTGCGCTGCTTCCACCGCAACACAGATGGCGGCATCAACG 829
   |||||
287 .....GlyLeu.....ThrAlaGlyLysSerAspProValIle 298

830 ATTGGCAGATATGGCGCACTCAACGACTATGCCGACGACCATCGC 879
   : : : : :
298 spVal...AspLeuSerGlnThrIleAspGlnGlnSerIle..... 310

880 GATTGGGCATCCAAACCCCAATGCCGACAGCATAGAACCGCTCAG 929
   |||||
311 .....ProGlnAlaGlnLysGlyPheTyrThrLysAs 321

930 CAATATCTTTAGCGCAGCATCATCCCGTCAAGGATGGAGCTGTTCGG 979
   : : : : :
321 rGlnGlnGlyThrAlaGlyLeuProPheAsp.....IleValS 334

980 GAATATACGCTTGCGCGCATCAGCGACATCTGTCAAGCGGTGCGAG 1029
   : : : : :
334 erAsnLeuAspLysLeuProAsnGlnAlaAsnAlaHisGlnLysSerArgSerLys 350

1030 ATGGCGCAGATCCGATTCGCGAAAGGAAATCCCGCTGACGAGCAATTT 1079
   : : : : :
351 LysLysHisThrGly...ProSerLeuSerSerAlaSerGlnProSerAl 366

1080 TGCGGATGGCGCATACGCCAAATACCTCCCTTACATTCGCGAAAT. 1128
   |||||
366 AlaIleSerSerSerSerSerSerLeuProSerAsnLeuAspLysIleAsn 383

1129 ..ATCGCTTCAACTGTGAG.....CAGGCTTAC 1155
   |||||
383 spValLysLysAsnIleGlnValSerAlaAsnGlnProGlnProArgPro 399

1156 GCGAAGGAACATCACTCACTCTCA.....ACGCT 1184
   |||||
400 ValLysLysAspValProLysSerGlnValGlyGlyLeuThrAspThr 416

1185 GCGCGCTCAAGGAAAGAAATGTGAATGCAACTGCAACAAACGCCACCGCA 1234
   : : : : :
416 rAspValIleAsnAsnSerThrProLysGlnGlnThrGlnGlnSerProS 433

1235 AGACCAAAAGTCCGCTTGACGAGTAAAGGTTTCCGATTTTGA AAAAGAC 1284
   : : : : :
433 erThrGlnLeuProGlnThrGlyLysGlnGlnProAsnLysAlaGln 449

1285 GTAAATATACGATACGAGATTAATACCGCTGTACCA...CAAGTGAATCC 1331
   |||||
450 ProAlaValProThrGlnAlaSerSerThrLysProSerGlnAlaAlaG 466

1332 TATATGATGAACCGCTTTTAAT.....CCTAAAGTTCTGTGGATCGG 1375
   : : : : :
466 uGlnSerThrProArgPheSerValArgProAsnLysPheThrGlySerA 483

1376 CTCATTTCTGG.....TCATTAACGTCAGAAATTCATACGCAAAATTA 1419
   : : : : :
483 rGlnAlaGlyPheValAlaAlaLeuGlnSerArgLeuGlnLysGlyProLeu 499

1420 CCAAGCGAAGGTAGATACATATATATCCACCTAAATAATTAATCTCTCTC 1469
   |||||
500 MetArgSer.....PheValProAsnLysSerLysSerProSe 512

1470 AGCAGCGCTACCAAAAGACCTAATATGATATTTGGATTAATAATTGTA 1519
   : : : : :
512 rGly.....ThrLysSerProAlaSerGlyGlnThrSerGlnAlaGly 527

```



```

1313 CTGTACCACAA.....GTGAA 1329
1086 roserprogluSerThrAspSerProgluArgspheleuproAan 1102
1330 CCTATGATGACCCGCTTTAATCTTAAGTCT..... 1365
1103 ProileuProProAlaSerLeuProGlnAsnSerThrleuProValTh 1119
1366 .....GTGGATCGGCTCATCTTGCTATACCTGACGACAAATTC 1405
1119 rGlnAgtgluValleuProIleHisSerGlnspIleSerLysProAlaP 1136
1406 AATACGCCAAATATTCACCA.....AGCGCAAGCTAGCAATC 1437
1136 roGlnProleuAlaProSerAspGlnCysLeuLeuArgGlnspThrVal 1152
1438 AGATATATCCACCTAAATATCTCTCTCAGCAGCCGCTACCAAAAG 1487
1153 GluThrThralaThrLeuSerThrProGlyVal...LeuProMetG1 1168
1488 ACCTAATAATGATATTGGATTAATTTGTAATGATGACTAA.... 1533
1168 yLysAspSerProleuLeuSerGlyArgGlyGluValValArgProL 1185
1534 .....GCTCA.....TCAAGAACTAAAGCTCAA 1557
1185 yAspAlaValAlaProAlaProleuLeuArgSerArgThrleuValLys 1201
1558 GAATTTGAATGGATGTTCA..... 1578
1202 ArgValThrTrpAsnLeuGlnGluAlaGlnAlaSerThrProAlaLeuAs 1218
1579 .....TTGCTAAACAGACAGACAGCAACTTGA. 1608
1218 parGAspProArgThrProleuGlnArgProGlnArgProGlnGlnGlyA 1235
1609 ..TGCGCTAGTAGGAT 1623
1235 sPTPraspAlaGluAsp 1240

seq_name: p1r2:S51342

seq_documentation_block:
verp10ln - yeast (Saccharomyces cerevisiae)
N:Alernate names: prol1n-rich protein VRP1; protein YLR337C
C:Species: Saccharomyces cerevisiae
C:Date: 23-Feb-1995 #sequence_revision 11-Aug-1995 #text_change 23-Mar-2001
C:Accession: S51342; S39626; S57435
R:Du, Z.
submitted to the EMBL Data Library, December 1994
A:Description: The sequence of S. cerevisiae cosmid 8300.
A:Reference number: S51339
A:Accession: S51342
A:Molecule type: DNA
A:Residues: 1-817 <D07>
A:Cross-references: EMBL:U19028; NID:9609380; PID:9609392; MIPS:YLR337C
R:Donnelly, S.F.H.; Pocklington, M.J.; Pallotta, D.; Orr, E.
Mol. Microbiol. 10, 585-596, 1993
A:Title: A proline-rich protein, verp10ln, involved in cytoskeletal organization and ce
A:Reference number: S39626; M01D:95035201
A:Accession: S39626
A:Molecule type: DNA
A:Status: nucleic acid sequence not shown
A:Residues: 1-307, 'R', 309-349, 'R', 351-688, 'E', 690-709, 'HLRWIPVPLIAPYKTLNNGYFLQVDRCTSTI
A:Cross-references: EMBL:Z26645; NID:9414785; PIDN:CA81388.1; PID:9439289
R:Munn, A.L.; Stevenson, B.J.; Gell, M.I.; Riezman, H.
submitted to the EMBL Data Library, June 1995
A:Description: ends, end6, and end7: mutations that cause actin delocalization and block
A:Reference number: S57435
A:Accession: S57435
A:Molecule type: DNA
A:Residues: 1-162, 'F', 164-817 <MUN>
A:Cross-references: EMBL:X87806; NID:9871534; PID:9871535

```

```

A:Experimental source: strain W303
A:Genetics:
A:Gene: SGD:VRP1; MDP5; ENDS
A:Cross-references: SGD:S0004329; MIPS:YLR337C
A:Map position: 12K
F:5-14/Region: proline-rich
F:77-85/Region: proline-rich
F:114-180/Region: proline-rich
F:216-245/Region: proline-rich
F:305-336/Region: proline-rich
F:349-357/Region: proline-rich
F:372-382/Region: proline-rich
F:396-406/Region: proline-rich
F:421-445/Region: proline-rich
F:518-528/Region: proline-rich
F:567-577/Region: proline-rich
F:608-621/Region: proline-rich
F:649-661/Region: proline-rich
F:678-685/Region: proline-rich
F:704-710/Region: proline-rich

alignment_scores:
Quality: 136.50 Length: 502
Ratio: 0.626 Gaps: 27
Percent Similarity: 43.426 Percent Identity: 24.502

alignment_block:
US-09-303-518D-465 x S51342 ..

Align seg 1/1 to: S51342 from: 1 to: 817

155 GCACGAGGGGGGAACTTCCGAGCGCA..... 181
15 AlaleuGlySerAlaProLysProAlaLysSerValMetGlnGlyAr 31
182 ....GCGCTCATATCGGATTTGGCAACATACAAAGCATCATGTTGGCAA 227
31 gAspAlaLeuLeuGlnLysPLeArgLysGlyMetLysLeuLysLysAlaG 48
228 CCTGTTCATCCAGCAGCGGCGCATTAAGAAATATGCGTACATTGTCC 277
48 luthrAsnAspArgSerAlaProIleValGlyGlyValValSerSer 64
278 GCTTTCCGATCAGGGGACG...AAGTCATTTCCCTTCGACAAACAT 324
65 AlaserGlySerSerGlyThrValSerSerLysGlyProSerMetSerAl 81
325 GCCTCACATTCCGATTTCTGATGAAGCGGTA..... 355
81 aProProIleProGlyMetGlyAlaProGlnLeuGlyAspIleLeuAlaG 98
356 .....GTCCCG.....TTGACGAGTATGCGCTTACCGCATCC 388
98 lylGlyIleProLysLeuLysHisIleAsnAsnAlaSerThrLysPro 114
389 ATTGAGGAGCATATACACCAATCCGCGACGCGCTATGACGGCCACAG 438
115 .....SerProSerAlaSerAlaProProIleProGlyAlaValProse 129
439 GCGCGCGCTATCCGCTCCCAAGGCGGAGGATATATACAGCTATCGA 488
129 rValAlaAlaProProIleProAsnAla.....ProLeuSerP 142
489 CATAAAGGCTTGCCTCAAAATATCCGCTTCACCTACCGACGACGCA 538
142 roAlaProAlaValProSerIleProSerSerAlaPro.....Pro 156
539 GCACCGGACAGCGGCTTGCGACGCTTCCACATACCGGAGTAGTACTG 588
157 lIleProAspIleProSerSerAlaAlaProProIleProIleVal..... 171
589 ACGCAAGAGTAGCGGACGAGTTCAAAGCGCGCACCC..... 625

```

```

172 .....ProSerSerProLeuProLeuProLeuSerg 183
626 ..GATACAGCCCGAGCTGG...ACAGATCGGGCAATGCCCGCAAGCT 670
183 lValSerAlaProLysValProGlnAsnArgProHisMetProSerVal 199
671 TCACGCGCAGCTGAGTATGTCMAAAACATCATCGCGCGGAGAGAGA 720
200 ArgProAla.....HisArgSerHisGlnArgly 209
721 ATTTGCGCGCAGCGGATGCCGTGAGGTATAGCGAAGGCTCAAAACAT 770
209 sSerSerAsnIleSerLeuProSer..... 217
771 TGCTGTATGACAGCGCTTGCTGCTTCCACCGAAACAGATGGCG 820
218 .....ValSerAlaProProLeuPro..... 224
821 GCATCAACGATTGGCAGATATGCGCAACTCAAGACTATGCCGAGCA 870
225 .....SerAlaSerLeuProThrHi 231
871 GCCATCCGCGATTGGGAGCTCCAAAACCCCAATGCCG..... 907
231 sValSerAsnProProGlnAlaProProProProProProThrIleG 248
908 .....CACAAAGCATGAAAGCCG.....TCAGCATATCTTTAGCG 943
248 lLeuAspSerLysAsnIleLysProThAspAsnAlaValSerProPro 264
944 CAGTCATCC...CCGTCAAGAGGATGTGAGCTGTTCGGGAAATAAGCGC 990
265 SerSerGlnValProLagLysLeuProPheLeuAlaGlnIleAsnAl 281
991 TTGGCGGCATCAGGCGACATC.....CTGTCAAGCGGTCCGAGAT 1031
281 aArgArgSerGlnArgLysAlaValAlaGlnLysValSerSerThrLysIleG 298
1032 GGGCGAGATCGCATTCGCGAAGGAAATCCGCG.....TCAGCG 1072
298 lThrIleAsnHisLysSerProSerGlnProProLeuProSerSerAla 314
1073 ACAATTTTCGCGATGCGGATAGCGCAATACCCGCTCCCTTACCATTC 1122
315 ProProIleProThrSerHisAlaProProLeuProProThrAlaProPr 331
1123 CGAAATATCGGTCAACTTTGGAGCGGTACGGCAAAAGAAACATCA. 1171
331 oProProSerLeuProAsnValThrSerAlaProLysLysAlaThrSerA 348
1172 .....CCACTCAACGCTGGCGCGCTCAACGGAAGAGA 1204
348 lAlaProAlaProProProProProProLeuProAlaAlaMetSerSerAlaSe 364
1205 ATGTGAAGACTGCAGAAACAAAGCGCA...CCCGAAGACCAAGTCCGTTT 1251
364 rThAsnSerValLysAlaThrProValProProThrLeuAlaPro... 379
1252 GAGCGTAAAGGCTTCCGAATTTGAAAAAGACGTAAATACGATACGAG 1301
380 .....ProLeuProAsn..... 383
1302 AATTAATACCGCTGTACCA.....CAAGTGAATCCATAGAG 1339
384 .....ThrThrSerValProProAsnLysAlaSerSerMetProAlaProP 399
1340 AACCCGCTTTAATCCTAAAGGTTCTGTGGATCGGCTATTCTTGCTCT 1389
399 roProProProProProGlnAlaPheSerThrSerAlaLeu 415
1390 ATTAAGTCCGAATTCATACGAAATTAACGAGGCAAGGTGATGATCAG 1439

```

```

416 SerAlaSerSerIleProLeuAlaProLeuProPro..... 428
1440 ATATATCCACCTTAAATAATTACTCT.....CCTTCAGACCGCTAC 1480
429 .....ProProSerSerValAlaThrSerSerValProSerAlaProProP 443
1481 CA 1482
443 ro 443
seq_name: p1r2:T42730

```

```

seq_documentation_block:
Bassoon protein - mouse
C:Species: Mus musculus (house mouse)
C>Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 21-Jul-2000
C:Accession: T42730
R:Dieck, S.; Sammarti-Vila, L.; Langnaese, K.; Richter, K.; Kindler, S.; Soyke, A.; W
J. Cell Biol. 142, 499-509, 1998
A:Title: Bassoon, a novel zinc-finger CAG/Glutamine-repeat protein selectively local1
A:Reference number: 222249; M0ID:98345363
A:Accession: T42730
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-3942 <DID>
A:Cross-references: EMBL:Y17034; NID:93413809; PIDN:CAAF6598.1; PID:93413810
A:Experimental source: strain 129 SVJ
C:Genetics:
A:Map position: 9F1
A:Introns: 72/2; 208/3; 505/3; 675/3; 2889/3; 3582/1; 3851/3; 3886/1; 3930/1
A>Note: bassoon
C:Function:
A:Description: may be involved in cytomatrix organization at the site of neurotransmi
A>Note: component of the presynaptic cytoskeleton
C:Keywords: coiled coil; zinc finger

```

```

alignment_scores:
Quality: 135.50 Length: 541
Ratio: 0.560 Gaps: 27
Percent Similarity: 44.732 Percent Identity: 22.551

```

```

alignment_block:
US-09-303-518D-465 x T42730 ..

```

```

Align seg 1/1 to: T42730 from: 1 to: 3942

```

```

95 TTATCCGCGAGCTTCTGACCGTCAGCATTTTCGAACCCGAGGGAATAC 144
||||| : : : : : ||| : : : : :
306 leuSerThrLysProSerThrAlaGlnProArgProProAlaGlnAl 322
145 CACGTAATCGGACGACGCGGGGAACTTCCGAGCGGACGCGTCATATCG 194
322 aglnLylsSerAla...ThrThrValProSerGlyLeuGlyAlaGly 338
338 lu...GlnThrGlnGlnLysLeuThrGlyLysLeuPheGlyAla 353
245 CGGCATTAAGGAATATCGGCTACATTCGCGCTTTCCGATCAGCGG 294
354 SerLeuLeuThrGlnAlaSerThrLeuMetSerValGlnProGlnAlaAs 370
295 CACGAAGTCATTCGCCCTTCGACACCAACCATCCACATTCGATTCGA 344
370 pThrGlnGlnInProSerProSerLysGlyPro..... 381
345 TGAAGCCGCGATGCCGTTGACGAGTTCAGCCTTTACCGCA..... 385
382 .....ProLysIleValPheSerAspAlaSerLysGlnAlaGlyProArg 396
366 .....TCCATTGGGAGGATACGACACCATTCGCG..... 415

```

```

397 ProProGlySerGlyProGlyProGlyProThrProGlyAlaLeuThrGlu 413
416 .CGAGCGGTATGACGGCCACAGGGCGGCTATCCGCTCCCAAGG 464
413 uProGlyAlaArgThrGlySerGlyProGlyAlaLeuAlaLeuAlaLeu 430
465 CGGAGGAGATATATACAGCTACGACATAAAGCGCTGCCAAATATCC 514
430 hteGlyGly.....GluHisGlnAlaLeuSerLysAla 432
515 GCCTCAACGTGACCGACACCGACCGGACGAGCTGTGCGACGCT 564
433 .....ThrAlaSerProLysHisGlyArgAla..... 441
565 TTCACAAATACCGGTAGATAGCTGACGCAAGAGTAGGCGAGGATTC 614
442 .....GluHisGlnAlaLeuSerLysAla 450
615 AGCGCGCACCGCATACAGCCCGGACGTGACAGATCGGCATGGCCG... 661
450 lAlaLeuProLysThrMetProLys.....GluArgAlaAlaLeuLys 464
662 CCGAAGCTTTCAACGGCACTGCAGATATGCTCAAAATCATCGCGCGG 711
465 ProLeuGlyGlnAlaGlnLeuAsnValGlySerArg...GlyProAlaAs 480
712 GGAGGAGAAATTTGCGCGCGAGCGATGCCCTGACGAGGTATAGCGAG 761
480 nTyraSerThrCysThrAla.....CysLysL 489
762 CTCAAACATGCTGTATGC..... 781
489 euGlnValCysAsnLeuGlyPheAsnProThrProHisLeuValGlu 505
782 ...ACGGCTTGGGTCTGCTTCCACCGAAACA...AGATGGCGGCGATC 825
506 LysThrGlnThrLeuPheCysLeuAsnGlyHisThrLysArgLeuGlnGlu 522
826 AACGATTTGGCAGATATGCGCGCACTCAAAAGTATGCCGACGAGCAT 875
522 ySerLeuGlyGlnProAlaProLeuProLeuProThrProGlnGlnProp 539
876 CGGCGATTGGCAGTCACAAACCCCAATGCCGACCAAGCATAGAGCGG 925
539 roAla.....GlyValProHisArgAla.....Ala 547
926 TCAGCAATATCTTTACGCGATCATCCCGTCAAGGATTTGACGCTGT 975
548 GlyAlaAlaProLeuLysGlnLysGlyProGlnGlyLeuGlyGlnProSe 564
976 CGGGGAAATACGGCTTGGCGGCGATCAGGCGACATCTGTCAACGGCTC 1025
564 rGlySerLeuProAlaLysAlaSerProGlnAlaThrLysAlaSer... 579
1026 CGAGATGGGCGAGATCGCATGCCGAAGGGAATCCCGCTGACGCGCA 1075
580 .....ProGlnAlaThrLysAlaSerProGlnAlaThrLys 591
1076 ATTTTGGCGATG.....CGGATACGGCAATACCGCTCCCTTACCAT 1119
592 AlaserProGlnThrThrLysAlaSerProGlnAlaLysProLeuArgAl 608
1120 TCCGCAATATCCGTTCAAACTTGAGCAGCGTTTACGCAAGAAACAT 1169
608 aThrGlnProSer...LysThrSerSerArgLysGlnLysLysThrA 624
1170 CACCTTCCTCAACCGTGGCGCGGTCAAGGAAAGATGTGAACCTGCGA 1219
624 lAthProAlaLysAla.....GluProValPr 633
1220 ACAAAAGCCACCGCAAGCAAGTCCGTTTGAACGTAAGGTTTCG 1269
633 OlyProProProGluThrThrValPro.....ProGlyThrPro 646

```

```

1270 AATTGTGAAAGACGTAAATACGATACGAGATTAATACCGCTTACC 1319
647 LysAlaLysSerGlyValLys.....ArgThrAspProAlaThrPr 660
1320 ACAAGTGAATCTTATAGATGACACCCGCTTTAATCTTAAAGGTTGTCG 1369
660 oValValLysProValProGlnAla.....ProLysGlyGlyGluA 674
1370 GA.....TCGGCT 1377
674 lArgGlnLysProValProLysProTyrSerGlnAspLeuSerArgSerPro 690
1378 CATCTTGGTCTATTAATGCGCAGATTCATACGCAAAATTAACAGGCA 1427
691 GlnSerLeuSerAspThrGlyTyrSerSerAspGlyValSerSerGly 707
1428 AGGTAGATCAATATATCCACCTAAATATAC.....TCTC 1465
707 nSerGlnLysThrGlyValAlaGlnGlnGlnValGlnGlnLeuAspSerA 724
1466 CTCGAGCAGCGGTACCAAAAGACCTAATATGATATTGATTAATTT 1515
724 lArgLysAlaThrGlyProArgProProSerProSerGlnLeuHisLysVal 740
1516 GGTAAAT..... 1521
741 GlySerSerLeuArgProSerLeuGlnAlaGlnAlaValAlaProSerAl 757
1522 .GAATGAGCTAAAGGTCCA 1539
757 aGluThrSerLysProPro 763
seq_name: p1r2:C39135
seq_documentation_block:
  hypothetical protein 3 (gyrB region) - Haloflex sp.
  C:Species: Haloflex sp.
  C:Date: 11-Feb-1993 #sequence_revision 11-Feb-1993 #text_change 18-Nov-1994
  C:Accession: C39135
  R:Holmes, M.L.; Dwyer, J.; Smith, M.L.
  J. Bacteriol. 173, 642-648, 1991
  A:Title: Mutations in DNA gyrase result in novobiocin resistance in halophilic archae
  A:Reference number: A39135; MUID:91100352
  A:Accession: C39135
  A:Status: preliminary; translation not shown
  A:Molecule type: DNA
  A:Residues: 1-437 <NOI>
  A:Cross-references: GB:M38373

```

```

alignment_scores:
  Quality: 135.00      Length: 466
  Ratio: 0.689        Gaps: 27
  Percent Similarity: 42.060      Percent Identity: 25.107

```

```

alignment_block:
  US-09-303-518D-465 x C39135 ..

```

```

Align seg 1/1 to: C39135 from: 1 to: 437

```

```

276 CCGCTTTCCGATCAGGGGACGAGTCCATTCCTCCCTGACAAACATG 325
111 111 111 111 111 111 111 111 111 111 111 111 111
21 ProArgThrArgArgArgHisArgAsnAspHisProLeu.....Le 34
326 CCGCAGATTCGATTCGATGAGAGCGGTAAGTCCGTTTACGAGTTTAC 375
111 111 111 111 111 111 111 111 111 111 111 111
34 uAlaGlyArgArgArgTyr.....LeuArgAspAspArgValArgLeuGlnA 49
376 CTTTACCGATCCATTTGGAGCGAGGTACGAACACATCCCGCGACGGCTA 425
111 111 111 111 111 111 111 111 111 111 111 111
49 sPalaArgLysProProAlaArgAlaArgValProGlnLeuArg..... 63

```

```

426 TCAGCGGCCACAGGCGG.....CGCTATCCGCTCCCAAGGCGCGA 469
      |||||
64 .....GlyArgSerPheAlaLeuArgIleAlaSerArgVal 76
      |||||
470 GGGATATATACAGCTACGACATAAAGCGCTGCCCAAAATATCCGCTC 519
      |||||
76 IGlnIleValProLeuArgGlyArgHisProArgValArgArgValPro 93
      |||||
520 AACCTGACCGCAACCCGACGCGGACAGCGGCTGTGACCGCTTCCA 569
      |||||
93 LArgSerPhe..... 96
      |||||
570 CAATACCCGCTAGTATCTGACGACGAGATAGCGACGATTCACACGCG 619
      |||||
97 .....AspGlyAlaProArgArg.....Arg 103
      |||||
620 CCACCCGATACAGCCCGA...GCTGACAGATGGCGCATGCCGCGCA 666
      |||||
103 GHisLeuLeuArgArgValIGlyGlnHisArgGlyArgSerArgHis 120
      |||||
667 GCTTCAAGCGGCACTG.....AGATATCGTCAAAACAT 701
      |||||
120 LArgIleArgArgAlaProGlyAlaSerArgLeuArgGlnHis 136
      |||||
702 ...CATGCGCGCGAGAGAAATTTGCGGCGACGCGATGCCGACAG 748
      |||||
137 GlnHisProArgGlyArgHis..... 143
      |||||
749 GTATAGCGAGAGCTCAACATTCCTTTATGACGCGTGGCTGCTT 798
      |||||
144 .....Ala 145
      |||||
799 TCACCGCAAAACAGATGCGCGCATCAAGATTTGCGATATGCGCGCA 848
      |||||
145 eArgArgValGlnSerPheAlaHisProArgArgGlnArgLeu 161
      |||||
849 ACTCAAGACTATGCG.....CGACGACCGCATCCGCGATGCG 886
      |||||
162 GlnProArgHisArgGlyArgProArgArgGlnProProArgGly 178
      |||||
887 C.....AGTCCAAACCC 900
      |||||
178 YArgSerArgGlyThrHisArgArgHisLeuArgGlnAlaProArgPro 195
      |||||
901 AATGCGCG.....ACAAG 914
      |||||
195 LValArgGlyProArgSerGlnAlaArgGlyPheArgGlyProArg 211
      |||||
915 CATAGACCGCTGACGAATATCTTTAGCGACGATCCCGCTCAAGGGA 964
      |||||
212 HisArgArgGlyArg.....His.ProProThrAla 222
      |||||
965 TTGAGAGCTGTGGGAAATACGGCTGGGCGGACATCAGCGCAATCCT 1014
      |||||
222 TArgSerAlaLeuArgGlyLeuProGlnHisGlyArgGlnHisLeu 237
      |||||
1015 GTCAAGCGGCTGACAGATGCGGAGATGCGCAAGAGGAAATCCGCG 1064
      |||||
238 ...GlnArgArgArg.GlyArgProArgProGlnArgGlnAlaGly 252
      |||||
1065 .....CGTACGCGCAATTTTGGCGGATGCGG 1090
      |||||
253 ArgGlyAlaHisProProGlnValArgAlaArgIleTyrLeuAlaAlaGly 269
      |||||
1091 C...ATAGCGCAATACCGCTCCCTTACGATTCGCGAAATATCCGCTCA 1137
      |||||
269 YGlnAlaArgGlyLeuProGlnProArgProLeuGlyValArgThrVal 286
      |||||
1138 AACTTGACGACGCTTACGCGCAAGAAACATACCTCCACACGCTCC 1187
      |||||
286 LArgGlyGlyArgLeuArgGlyArgValGlnAlaGlyProArgPro 302
      |||||
1188 GCGCTC.....AACGGAAGAGATGTGAATCTGGCAACA 1222

```

```

      |||||
303 GlnValProGlyArgPheAlaProGlnGlyGlnSerGlyArgArgGly 319
      |||||
1223 AACGCGCCGAGACGACCAAGTCCGCTTACGCGTAAAGGCTTCGAT 1272
      |||||
319 YThrProProArgProHisSerArgGlyArgArgSerPheGlyAlaHis 336
      |||||
1273 TT.....TGAAAGAGCTAAATACGATACGATTAATATAC 1310
      |||||
336 LArgHisThrArgArgArgArgArgArgValArgHis.Arg...Gln 351
      |||||
1311 CGCTGACCAAGTAATCTTATAGTAAACCGCTTATCTTAATCTTAAG 1360
      |||||
351 YAlaLeuProAlaAlaHisProArgSerArgArgArgArgArgG 368
      |||||
1361 GTTCTGCGATGCGCTCATCTTGGCTATATCTGCGCAATTCATATAC 1410
      |||||
368 LHisProSerAlaAla.....AlaTyr 375
      |||||
1411 GCAAAATTACCA.....AGCGAAGTATACATCAGATATAT 1445
      |||||
376 LAlaSerValProAlaHisAlaProAlaHisArgGlyArgLeuArg...Va 391
      |||||
1446 CCCACCTAAATATCTCTCTTACGACCGCTTACCAAA 1485
      |||||
391 LArgGlySerThrAlaAlaValAlaProArgProLeuProArg 404
      |||||
seq_name: p1r2:G01763
seq_documentation_block:
atrophin-1 - human
C:Species: Homo sapiens (man)
C:Date: 21-Dec-1996 #sequence_revision 06-Jun-1997 #text_change 17-Jul-1998
C:Accession: G01763
R:Margolis, R.L.
submitted to the EMBL Data Library, March 1995
A:Reference number: G08343
A:Accession: G01763
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-1184 <MAP>
A:Cross-references: EMBL:U23851; NID:915325; PID:915326
C:Genetics:
A:Gene: GDB:DRPLA; B37
A:Cross-references: GDB:270336; OMIM:125370
A:Map position: 12p-12p
alignment_scores:
      Quality: 134.00      Length: 682
      Ratio: 0.515      Gaps: 35
Percent Similarity: 38.123      Percent Identity: 22.141
alignment_block:
US-09-303-518D-465 x G01763 ..
Align seg 1/1 to: G01763 from: 1 to: 1184
65 CACGCGCTCAGATTTGCAACGATCTTTATCCGCGAGTTCTCGAC 114
      |||||
252 HisProPro.....ProThrThrProIleSerValSerSerGly 265
      |||||
115 CGTACGATTTGCAACCGCGACGCGAAATACCACTATTCGCGAGGG 164
      |||||
265 YAlaSerGlyAlaProProThrThrLysProProThrThrProValGly 282
      |||||
165 GGAACCTTGGCGAGCGCGCTCATATCGATTTGGCAACATCAAGCC 214
      |||||
282 LysAlaLeuProSerAlaProProProAlaAspHe.ProHisValThrPr 298
      |||||
215 ATCAGTTGGGCAACCTTTCATCCAGCAGG.CGGCAATTAAGAAATAT 263
      |||||
298 OAsnLeuProProPro.....ProAlaLeuArgProLeuAsnAla 313

```

```
264 CGGCTACATGTCGGCTTTCCGATCCGCGCAGCAAGTCATCCCT 313
|||||
313 erAlaserPro.....ProGlyLeuGlyAlaGlnProLeuProGly 326
314 TCGACACACATGCTTCACATTCGGATTTCGATGAAGCCGGATGCCGTT 363
|||||
327 HisLeuProSerProHis..... 332
364 GACGATTCAGCCTTTACCCGATCCATGGACGCGATACAGACACATCC 413
|||||
333 .....AlaMetGlyGlnGlyIleGlyGlyLeuP 342
414 CGCGCAGGCTATGACGGCCACAGGGCGCGCTATCCG.....CTC 457
|||||
342 roProGlyProGlyGlyProThrLeuAlaProSerProHisSerLeu 358
458 CCAAGGCCGCGAGGATATATACAGTACGACATATAAGGCTGCCCA 507
|||||
359 ProProAlaSerSerSerAlaProAlaProPro..MetArgProThy 374
508 AATATCCGCTCAACCTGACCGCAACAGCAGCAGCGCAACGGCTGT 557
|||||
374 rSerSerSerSerSerSerAlaAlaAlaSerSerSerSerSers 391
558 CGACGCTTTCCAAATACCGGTAGTATGCTGACGCAAGAGTAGCGCAG 607
|||||
391 erSerSerSerAlaSerProPhe..... 398
608 GATTCAAGACCCGCCACCGCATACAGC...CCGACGTGACGATCGGCG 654
|||||
399 .....ProAlaSerGlnAlaLeuProSerThyProHisSerPhe 411
655 AATGCCCGCAAGCTTTCAACGCGACTCGATATCGTCAAAAACATCAT 704
|||||
411 erProPro.....TherSerLeuSerValSerInGlnProPol 425
705 CGCGCGCGCAGAGAAATTGTCGCGCAGAGCGATGCCGTGACGGTATA 754
:::
425 yArThrGlnProSerLeuProSerGlnAlaVal..... 436
755 GCGAAGGCTCAAAACATGCTTTATGACAGGCTGGGTGCTGTTCCACC 804
|||||
437 .....TrpSerGlnGlyProP 442
805 GAAAAA.....AGATGGCGCGCATCAACGATTTGGCAGATAT 842
|||||
442 orProProProGlyArgLeuLeuAlaSer..... 454
843 GCGCGACATCAAGACTATGCGCGCAGCAGCCATCGCGCATGGCAGATCC 892
|||||
455 .....AsnAlaHisProGlyProPheProProSerThrGlyAlaGlnSer 469
893 AAAACCCCATGCG.....CACAAAGCATAGAAGCGGTACG 930
:::
470 ThrAlaHisProProValSerThrHisHis..HisHisHisGlnGlnGln 486
931 AATATCTTTACGCGCAGTCATCCCGTCAAGGAGATTGAGCTGTTCGGG 980
|||||
486 lnglnGlnGlnGlnGlnGlnGlnGlnGlnGlnHisHisGlyAsnSerGly 502
981 AAAATACGGCTTGG.....CGGCA 1000
|||||
503 ProProProProGlyAlaPheProHisProLeuGlnGlySerSerHis 519
1001 TCACGCGACATGCT..... 1014
|||||
519 HisHisProGlyAlaMetSerProSerLeuGlySerLeuArgPro 535
1015 .....GTCAAGCG 1022
|||||
536 TyrProProGlyProAlaHisLeuProProProHisSerGlnValSerTy 552

1023 GTCCGAGATGGCGCAGATCCGATTCGCCAAAGGAATCCGCC...GTCA 1069
|||||
552 rSerGlnAlaGlnProAsnGlyProProValSerSerSerSerHisSers 569
1070 GCGCAATTTTGGCGATGCCGATAC..... 1095
|||||
569 erSerSerThrSerGlnGlySerTyProCysSerHisProSerProSer 585
1096 .....GCCAATACCGGCCCTTACCATTCGCCGAATA 1129
|||||
586 GlnGlyProGlnGlyAlaProTyProPhePro..ProValProThyVal 602
1130 .....TCCGTTCAACTTGGACGCGCTTACGCA..... 1159
|||||
602 hrThrSerSerAlaThrLeuSerThrValIleAlaThrValAlaSerSer 618
1160 .....AAGAAACATCACCTCTCTCAACCGTGGCGCGGTCAAC 1197
|||||
619 ProAlaGlyTyThrThrAlaSerProProGlyPro..ProPro..... 632
1198 GGAAGAATGTGAACACTGGCAACAAACGCCACCG..... 1233
633 .....TyGlyLysAlaGlnAlaProSerProGlyAlaTy 643
1234 .AAGACCAAGATGCCGTTTGACGTTAAGGTT.....C 1267
|||||
643 rLysThrAlaThrPro.....ProGlyTyThrLysProGlySerProP 657
1268 CGAATTTTGAAGAAAGAGTAAATACGATACGGA..... 1302
|||||
657 roSerPheArgThrGlyThrProProGlyTyThrArgGlyThrSerProPro 673
1303 .....ATTATACCGCTGTACCAAGTGAATGATCTTATGA 1337
674 AlaGlyProGlyThrPheLysProGlySerProThyValGlyProGlyThr 690
1338 TGAACCGCTTTAATCTTAAGT..... 1362
|||||
690 oleProProAlaGlyProSerGlyLeuProSerLeuProProProPro 707
707 laAlaProAlaSerGlyProProLeuSerAlaThr..... 718
1411 GCAAAATTACCAAGGCAAGGTAGATCAGATATATCCACTAAATAATTA 1460
|||||
719 .....GlnIleLysGlnGlnProAlaGlnGlyTy 728
1461 C...TCTCCTTACGACACCGCTACCA.....AAGGACCTAAT..... 1494
|||||
728 rgIuThrProGlySerProValProProAlaArgSerProSerProProP 745
1495 .....AAT 1497
745 roLysValAlaPheValProSerHisAlaSerGlnSerAlaArgPheAsn 761
1498 GGATATTGGATAAATTGGTAAT..... 1521
|||||
762 LysHisLeuAspArgGlyPheAsnSerCysAlaArgSerAspLeuTyThr 778
1521 ..... 1521
778 eValProLeuGlnGlySerLysLeuAlaLysLysArgAlaAspLeuValG 795
1521 ..... 1521
795 LuLysValArgArgGlyAlaGlnGlnArgAlaArgGlnGlyLysGlyArg 811
1522 .....GAATGACATAAGGCTTCATCAAGACTAAAGGTCAAGA 1559
|||||
812 GluArgGluArgGluArgGluLysGluArgGluArgGluLysGluArgGlu 828
1560 ATTTGAATGGGATGTTCAATTGCTTAACACAGGAGA 1596
```

828 uleuGlunrGserValLysLeuAlaGlnGlyArg 840
seq_name: plr2:T33369

seq_documentation_block:
hypotheoretical protein H02F09.3 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 29-Oct-1999 #sequence_rev: 100 29-Oct-1999 #text_change: 17-Mar-2000
R:Geisel, C.; Harmon, G.
Submitted to the EMBL Data Library, July 1998
A:Description: The sequence of C. elegans cosmid H02F09.
A:Reference number: Z21330
A:Accession: T33369
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1275 <GEN>
A:Cross-references: EMBL:AF077538; PDB:1AC64622.1; GSPDB:GN00028; CESP:H02F09.3
A:Experimental source: strain Bristol N2; clone H02F09
A:Gene: CESP:H02F09.3
A:Map position: X
A:Introns: 42/1; 78/3; 106/3; 135/2; 161/2; 224/1
C:Superfamily: yeast glucan 1,4-alpha-glucosidase homolog; glucan 1,4-alpha-glucosidase

alignment_scores:
Quality: 136.00 Length: 458
Ratio: 0.642 Gaps: 20
Percent Similarity: 46.288 Percent Identity: 22.052

alignment_block:
US-09-303-518D-465 x T33369 ..

Align seg 1/1 to: T33369 from: 1 to: 1275

```

92 CTTTATCCGCGAGGTTCTGACGCGATTCG...AACCGACGG 138
   |||||
177 ValIleThrLysThrPheValIleValSerLeuSerLeuAsnSerThrAs 193
   |||||
139 A...AATACCACTATTCGCGACGAGGGGGAATCTGCGACGCGACGG 185
   |||||
193 pmetAsnSerArgTyrGlyLysAlaIleAsnIleProThr 208
   |||||
186 TCATATCGGATTGGGAACATACAAAGCATGAGGCGAACCTGTTC 235,
   |||||
208 ..... 208
236 TCCAGACGCGCGCATTAAGAAATATGCGTACATGTCGCGTTTCC 285
   |||||
209 .....ProthrGlnSpliSerAsnLysIleAsnIleLe 221
286 GATCAGCGGACGAGTCCATTCCTCCCTTCGACAAAC... 322
   |||||
221 uasnlIleGlyThrThrGlnThrProValThrThrSerThrMetAla 238
   |||||
232 .....ATGCCCTCAC.....ATCCGAT 340
238 hrThrThrAlaAsnValThrSerAlaAlaProAsnThrThrValThrIle 254
   |||||
255 SerThrSerProThrThrValThrValProSerThrAlaGlnThrSe 271
   |||||
391 TGGGACGATACGACACCATCCCGCGACGCGTATGACGGCGACAGG 440
   |||||
271 rSerThrThrThrValThrValProThrThrThrValThrGly...Prot 287
441 CGGCGGCTATCCGCTCCCAAGGCGGACGATATATACGACGACGCA 490
   |||||
287 hrThrValThrValProThr.....ThrValThr 298

```

```

491 TAAAGCGTTCCCAAAATATCCGCTCACTGACGCAAC... 535
   |||||
299 IleProSerThrValThrSerProIleThrThrProSerThrValVa 315
   |||||
536 .GCAGCAGCGGACACGCGCTTGTGACCGCTTCCAAATACGGTGAT 584
   |||||
315 lThrValPro.....SerThrValThrValProSerThrVa 328
   |||||
585 GCTGACGCAAGAGTAGCGACGATTCAAACGCGCCACCGCAT... 628
   |||||
328 lAlaThrLysPro...SerThrValThrAlaProSerThrValVal 343
   |||||
629 .....ACAGCCCGAGCTGACGATCGGCGAT 657
   |||||
344 ThrValProSerThrValThrLysProAsnThrValThrValThrSe 360
   |||||
658 GCCCGGACGCTTTCACGCGCATGATGCAATGCAAAACATCATCGG 707
   |||||
360 rProThrValAlaThrThrProThrThrValThrThrProSerThrV 377
   |||||
708 CGGCGCAGAGAAATGTGCGCGACGCGATCCGTGACGGTATAAGCG 757
   |||||
377 aThrValThrValProSerThrValThrValProThrThrValThr 393
   |||||
758 AAGGCTCAACATTCGCTGTATGACGCGCTGCTGCTTCCACCGCA 807
   |||||
394 AsnProSerThrValThrAlaProSerThrThrValThrValProTh 410
   |||||
808 AACAGATGCGCGCA...TCACGATTTGGCAGATATGCGCAGCACTCA 854
   |||||
410 rThrValMetThrSerArgSerThrValIleThrThrProThrThrG 427
   |||||
855 AGACTATCGCGCAG..... 868
   |||||
427 LysSerProSerThrAlaGlyThrSerLeuAlaSerThrAlaValThr 443
   |||||
869 CAGCATCCGCGATTTGGCGAGTCCAAACCCCAATGCCGCAAGGCTA 918
   |||||
444 ThrGlnThrSerIleGlySerSerSerThrProLeuProSerGlnSer 460
   |||||
919 GAAGCGTCAGCAATATCTTACGCGACTCATCCCGTCAAGGATTTGG 968
   |||||
460 rSerLeuSerMetSerSerLeuSerThrThrThrProSer..... 473
   |||||
969 AGCTGTTCGCGAAATATGCGGCGGCGATACGCGACATC..... 1012
   |||||
474 .....SerSerThrAlaGlyAlaThrSerProAlaThrGlnGln 486
   |||||
1013 .....CTGTCAAGCGGTGCGACGATGG.. 1033
   |||||
487 SerThrLysProThrIleGlyThrSerMetSerSerGlyProThrThr 503
   |||||
1034 .....GCGAGATGCGATTTGCCGAAAGGAAATCCGCGT 1067
   |||||
503 lAlaProGlyAlaSerThrGlnSerThrValLeuGlnSerSerThrPro 520
   |||||
1068 CAGGACAAATTTGCCGATG.....CCGCAATACGCGCAAT 1102
   |||||
520 erglyThrThrValThrLeuProSerGlySerSerThrAlaThrAlaGly 536
   |||||
1103 ACCGCTGCC.....CTTACGATTCGCGAAATATCCGTTCA 1137
   |||||
537 ThrSerProGlnAlaSerThrValThrValThrAspIleSerThrVa 553
   |||||
1138 AACTTGAGCAGCGTTACGGCAAA.....GAAACATACCTCTCTCA 1180
   |||||
553 lSerGlySerThrValThrSerGlnThrAlaGlnSerSerLeuSerThr 570
   |||||
1181 CGGTCCGCGCGTCAAAAGGAAAGATGTGAACCTGCAAC.....AAA 1224
   |||||
570 lSerProThrSerAlaGlySerSerIleSerThrValSerThrValSer 586
   |||||
1225 CGCGACCGGAAAGCAAGTGGCCG 1248

```


587 SerGlnProSerThrTyrIlePro 594

seq_name: p1r1:EDBEIF

seq_documentation_block:
immediate-early protein IE180 - suid herpesvirus 1 (strain Indiana-Punkhauser)

C:Species: suid herpesvirus 1

C:Date: 30-Jun-1990 #sequence_revision 30-Jun-1990 #text_change 16-Feb-1997

C:Accession: S04713

R:Cheung A.K.

Nucleic Acids Res. 17, 4637-4646, 1989

A:Title: DNA nucleotide sequence analysis of the immediate-early gene of pseudorabies vi

A:Reference number: S04713; M01D:89315207

A:Accession: S04713

A:Molecule type: DNA

A:Residues: 1-1460 <CHE>

C:Superfamily: herpesvirus immediate-early protein IE175

C:Keywords: DNA binding; early protein; transcription regulation

alignment_scores:

Quality: 134.00 Length: 607
Ratio: 0.575 Gaps: 30

Percent Similarity: 38.386 Percent Identity: 22.570

alignment_block:

us-09-303-518d-465 x EDBEIF ...

Align seg 1/1 to: EDBEIF from: 1 to: 1460

```

117 TCAGCATTCGACACCGCGGAAATACCACTATTCGCGACGAGGGGG 166
    ||||| ||||| ||||| |||||
75 AlaAlaGlyAlaThrArgProProArgProProSerAlaGlnGln... 89
167 AACTGCCGAGCGCGCGTCATATCGATTGG... 200
    ||||| ||||| |||||
90 .....GlnArgHisAlaArgArgGlySerGlyGluIleVal 102
201 .....AAACATACAAAGCCATTCAGTT 221
102 AlLeuAspSerGluAspGluGluAspGluProGlySerProAlaIa 118
222 GCGCAACCTGTT... 233
    ||||| ||||| |||||
119 GlySerProValGlyLeuSerIleArgAlaProSerThrValThrSer 135
234 .....CATCCAGCAGGGGCCATTAAGGAAATATCGGC 267
135 rSerGlyProGlyProGlyProAlaProGlyProGlyArgProAla 152
268 TACATTGT... 296
    ||||| ||||| |||||
152 LmHisSerIleArgGlnArgProGlyProProAlaIaProGlyAla 168
297 CGAAGTCACATTCCTTCGACACACATGCCATTCGATTCGATG 346
169 ProProGluProGluProProArgProPro... 177
347 AAGCGGTGTCCTTCGACGATTCAGCTTACCGCATTCATGGGAC 396
178 .....ProProAlaProProAla 185
397 GGATACGAACACCATCCGCCGACGG... 422
185 roProAlaProProAlaProAlaArgProArgGlyAspGlyProPro 201
423 .....CTATGAGGGCGACAGGGCGCATTCGCCGCCCAAG 463
202 GlyGlyThrArgSerValSerProGlyArgArgGlyLeuGlyPro 218
464 GCGCGAGGATATACAGCTACGACATTAAGGCGT... 500
    ||||| ||||| |||||

```

```

218 gArgHisGlnHisSerGlnGlnArgTrpProGlnArgGlnGlyGly 235
501 .....TGCCCAAAATATCCGCTCAACCTGA... 532
    ||||| ||||| |||||
235 LProLeuProGluProProProProProGlyArgSerArgProAla 251
533 ACCGCGACCGCGACAAACGCTTCGAC...GTTCACAAATACCGGT 579
    ||||| ||||| |||||
252 AlaAlaAlaProProProAlaGlnGlyThrAlaValThrIle... 266
580 ACTATGCTGACGCAAGAGTAGGCGAGGATTCAAACCGCCACCC... 625
    ||||| ||||| |||||
267 .....ThrSerThrAlaSerProTrp 274
626 .....GATACAGCCCGAGCTGACGA 646
    ||||| ||||| |||||
274 euAspGluProAlaIaAlaIaArgAlaAspProAlaIaAlaTrp... 289
647 GATCGGCAATGCGCCGACGCTTCACACGCGCATCGAGATTCGTCAA 696
    ||||| ||||| |||||
290 ...ArgProGluProArgLeuGlnProGlnLeuGlnLeu 305
697 AACATCATCGCGCGGCGAG... 716
    ||||| ||||| |||||
305 LSHHisArgArgArgAlaArgArgProArgProArgGluGlyArg 321
717 .....AGAAATG 724
322 GlyArgThrArgProArgArgGlyArgGlyAlaProLeuGlnArg 338
725 TCGGCGCAGCGCATGCGCGCATTAAGGAGGTCAAAATTCCT 774
    ||||| ||||| |||||
338 oArgArgArgArgAlaGlyGlyGlyAlaLeuArg... 350
775 GTTATCAGCGCTTGCTGCTTCACCGAAACAGATGCGCGCAT 824
    ||||| ||||| |||||
351 .....GlyArgGlyPheSerSerSerSerSerGly... 360
825 CAACGATTTGGCAGATATGCGCGCACTCAAAAGACT...TGCGCGAG 868
    ||||| ||||| |||||
361 .....GlySerAspSerAspLeuSerProAlaArgSe 371
869 C...AGCCATCCGCGATTGGCGATCCAAACCCCAATCCGCGACAGGC 915
    ||||| ||||| |||||
371 rProSerAlaProArg...AlaProAlaIaAlaIaAlaIaAla 386
916 ATAGAACCCTCAGCAATATCTTACGCGAGTCATCCGCTCAAGGAT 965
    ||||| ||||| |||||
386 rGlySerAlaSerSerSerSerSerSerSerSerSerSerSer 402
966 TGGAGCTGTTCGGGAAATACGCTTGGCGGCGATCAGCGCATCTG 1015
    ||||| ||||| |||||
403 SerSerSerGluGlyGluAspGlyValArgProGlyAla 419
1016 TCAAGCG... 1050
    ||||| ||||| |||||
419 euAlaIaArgAlaGlyProProProSerProProAlaIaAlaIa 435
1051 AAA...GGGAAATCCGCGCGACGACATTTTGGCGATCGGACATCGC 1097
    ||||| ||||| |||||
436 ArgProSerAlaSerSerAlaSerAlaThrSerSerAlaIaAla 452
1148 AGCGTTACGCGCAAGAAACATCACTCCCAACCGCGCGCTCAAC 1197
    ||||| ||||| |||||
469 erThrAsnAsnHisLeuSerLeuMetAlaAspGlyProProTrpHis 485
1198 GGAAGAAGATGTGAACCTGGCAACCAACGCCCGCGAAGCAAGTGCC 1247
    ||||| ||||| |||||
486 GlyProLeuLeuThrProLeuGlyGluProTrpProGlySerAspPro 502
    ||||| ||||| |||||

```


C:Species: Homo sapiens (man)
 C:Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 20-Apr-2000
 C:Accession: T47182
 R:Koehler, K.; Beyer, A.; Mewes, H.W.; Weil, B.; Wiemann, S.
 Submitted to the Protein Sequence Database, March 2000
 A:Reference number: Z24378
 A:Accession: T47182
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-1299 <AAA>
 A:Cross-references: EMBL:AL162004
 A:Experimental source: adult testis; clone DKFZp434M1616
 C:Genetics:
 A>Note: DKFZp434M1616.1

alignment_scores:
 Quality: 132.50 Length: 567
 Ratio: 0.465 Gaps: 31
 Percent Similarity: 50.265 Percent Identity: 22.046

alignment_block:

US-09-303-518D-465 x T47182

Align seg 1/1 to: T47182 from: 1 to: 1299

```

74 CAGATTGG.....CAACGAT 90
   |||:||||
254 GluValTrpAsnLysAsnAlaAsnGluLysGlyArgSerGlnThr 270
   |||:||||
91 TCCTTTTCGGCGCGCTTCGACCGCAGACATTTGCAACCCGACGGAA 140
   |||:|||||
270 LysLeuProProLysPheAlaLys.....LysGlnAlaThrGly 284
   |||:|||||
141 ATACACCTTATTCGGCAGCAGGGGGAACTTCCGAGCGCGGTCATA 190
   |||:|||||
284 LeuGlnAlaGlnSerSerAlaSerValProProLeuAlaSerAlaPro 300
   |||:|||||
191 TCGGATGGGAAATATCAAGCCATCAGTGGGCAACCTGTTCAATCAG 240
   |||:|||||
301 LeuProProSerThrSerAlaSerValPro.....AlaSerTh 313
   |||:|||||
241 CAGCGCGCCATTAAGAAATATCGCTACATGTCGCGTTTCGATCA 290
   |||:|||||
313 SerAlaProLeu.....ProAlaThrLeuThrProValProAla 327
   |||:|||||
291 CGGCGCAGAAATCCATTCCTCCACACCATGCTTCACAT 334
   |||:|||||
327 ThrSerAlaProValProAlaSerThrLeuAlaProValLeuAlaSer 343
   |||:|||||
335 .....CCGATTCGATGAAGCCGCTAGCCGTTGACGATTCAGC 375
   |||:|||||
344 ThrSerAlaProValProAlaSerProLeuAlaProValSerAlaSer 360
   |||:|||||
376 C...TTTACCGCATTCATGGGAGGATAGCAACACCATCCGCGCAGG 422
   |||:|||||
360 SerValSerAlaSerValProAlaSerThrSerAlaAlaLeuThrS 377
   |||:|||||
423 CTATCAGGGCCACAGGGGGGCGATCCCGTCCCAAGCGCGAGGG 472
   |||:|||||
377 ThrSerSerProAlaSerAlaProAlaProThrProLeuAlaSer 393
   |||:|||||
473 ATATATACA..... 481
   |||:|||||
394 ValSerThrProAlaSerValThrLeuAlaSerAlaSerLeuPro 410
   |||:|||||
482 .....GCTA 485
   |||:|||||
410 LeuAlaSerAlaLeuAlaSerThrSerAlaProThrProAlaProAla 427
   |||:|||||
486 CGACATTAAGCGCTTCCCAAAATATCCGCTCAACCGGACCGCAGC 535
   |||:|||||
427 LysSerSerProAlaAlaProValIleThrAlaProThrIleProAlaSer 443

```

```

536 GCA.....GCACGGCAGACAGCGCTTCGACCG 564
   |||:|||||
444 AlaProThrAlaSerValProLeuAlaProAlaSerAlaSerAlaProAl 460
   |||:|||||
565 T...TCCCAATATCCGCTAGTATGCTGACGCAAGAGATAGCGACGAT 611
   |||:|||||
460 AlaProAlaProThrProVal.....SerAlaProAla 470
   |||:|||||
612 CAACCGCGCCCGCATACAGCCCGAGCTGACAGATCGGCAATGCCG 661
   |||:|||||
470 snProAlaProPro.....AlaProAlaGlnThrGlnAlaGlnThrHis 484
   |||:|||||
662 CCGAAGCTTTCACAGCGCAGTATGATGTCATAAACAATCCGCGCG 711
   |||:|||||
485 LysProAlaGlnAsnProLeuGlnThrThrSerGlnSerSerLysGln 501
   |||:|||||
712 GCAGGAGAAATG.....TCGGCGCAGGC...GATCCCGTGCAG 748
   |||:|||||
501 opProProSerThrLeuArgLeuProSerAlaGlnThrProAsnGlyThrAsp 518
   |||:|||||
749 GTATAAGCGAAGCTCAACATTTGCTATGACAGCGCTTGGCTGCT 798
   |||:|||||
518 yValAlaSerGlyLysSerIleGlnThrProGlnSerHisGlyThrLeu 534
   |||:|||||
799 TCACCGCA.....ACAAGATGCGCGC.....ATCAGCA 830
   |||:|||||
535 ThrAlaGlnLeuTrpAspAsnLysValAlaProProAlaValLeuAsn 551
   |||:|||||
831 TTTGGCAGATATGGCGCACTCAAGACTATGCGCAGCAGCATCCCG 880
   |||:|||||
551 pLysSerLysLysLeu.GlyProLysProProGlnProProSerVal 567
   |||:|||||
881 ATTGGCAGCTCCAAACCCCAATGCCGCAACAAGCATGAAGCCGTCAG 930
   |||:|||||
568 SerAlaTrpAsnLys...ProLeuThrSerPheGlySerAlaProSerSe 583
   |||:|||||
931 AATATCTTTACGCGCATGATCCCGTCAAGAGGATTTGAGCTGTTCCGG 980
   |||:|||||
583 rGluGlnAlaLysAsnGlyGlnLysSer...GlyLeuIle...GlyT 598
   |||:|||||
981 AAAATACG.....GCTTGGCGGCGATCAGCGCAGCATCTGCA 1018
   |||:|||||
598 hrAspThrIleGlnPheGlyAlaProAlaSer..AsnGlyAsnGluAsnG 614
   |||:|||||
1019 AGCGGTC.....GCAGTGGCGGAGATTCGCTTCGCAAAAGG 1056
   |||:|||||
614 uValValProValLeuSerGluLysSerAlaAspLysIleProGluPro 631
   |||:|||||
1057 AAATCCGCGCTCAGGACATTTTGGCCGATCGGC.....ATACGC 1097
   |||:|||||
631 ySglnGlnArgLysGln...ProAlaGlnGlyProLysAlaGln 646
   |||:|||||
1098 CAATATCCCGTCCCTTACCATTTCCCAATATTCGTTCAAACTGGAGC 1147
   |||:|||||
647 LysLeuProAspLeuSerProValGlnAsnLysGlnHisLysProGlyPr 663
   |||:|||||
1148 AGCGTTACGGCAAGAAACATCACTCTCCACCGCGCGCTCAAC 1197
   |||:|||||
663 oLleGlyLysGlnArg.....SerLeuYSA 672
   |||:|||||
1198 GGAAGAATGTGAAGACTGGCAAAACGCCACCGAAGCAAGT... 1244
   |||:|||||
672 snArgLysValLysAspAlaGlnValGluProGlnGlyGlnGluLys 688
   |||:|||||
1245 .....GCCGTTGACGTTAAGGTTTCGCAATTTGAAAGACG 1285
   |||:|||||
689 ProSerProAlaThrValArg.SerThrAspProValThrThrLysGlu 705
   |||:|||||
1286 TAAATATGATACGAGATTAATACCGCTTACACAAAGTAATCTGTATA 1335
   |||:|||||
705 hrLysAlaValSerGlnMetSerThrGlnIleGlyThrMetIleSerVal 721

```

```

1336 GATGACCCCTCTTATCTTAAGTTCTGCGATCGGCTCAT...TC 1382
      :::::::::::::::::::: |||||:::
722 SerSerAlaGlyThrAsnAlaLysMetGluSerAlaArgLysAl 738
      :::::::::::::::::::: |||||:::
1383 TTGGTCATTAACGCCGAATTCATACGCAAAATACCAAGCAAGGA 1432
      :::::::::::::::::::: |||||:::
738 atpGluAsnSerProAsnValArgGluLysGlySerProValThrSer 755
      :::::::::::::::::::: |||||:::
1433 GAATCAGATATATCCACCT.....AAAATTAATCTCTCTGACGACCG 1476
      :::::::::::::::::::: |||||:::
755 hr.....AlaProProlLeuAlaThrGlyValSerSerSerAla... 767
      :::::::::::::::::::: |||||:::
1477 CTACCAAAAGACCTAATATGATATTGGATTAATTGGTAAT 1521
      :::::::::::::::::::: |||||:::
768 .....SerGlyProSerThrAlaAsnTyrAsnSerPheSerSer 780
      :::::::::::::::::::: |||||:::

```

seq_name: p1r2:113606

```

seq_documentation_block:
  hypothetical protein 87B1.5 - fruit fly (Drosophila melanogaster)
  C:Species: Drosophila melanogaster
  C:date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 17-Nov-2000
  C:Accession: T13606; S23632
  R:Murphy, L.; Harris, D.; Barrell, B.
  Submitted to the EMBL Data Library, April 1999
  A:Description: Sequencing the distal X chromosome of Drosophila melanogaster.
  A:Reference number: Z17668
  A:Accession: T13606
  A:Status: preliminary; translated from GB/EMBL/DBJ
  A:Molecule type: DNA
  A:Residues: 1-1589 <MUR>
  A:Cross-references: EMBL:Z98269; NID:e1355202; PID:e1251078; PIDN:CAB10975.1
  R:Decamillis, M.; Cheng, N.; Pierre, D.; Brock, H.W.
  Genes Dev. 6, 223-232, 1992
  A:title: The polyhomeotic gene of Drosophila encodes a chromatin protein that shares pol
  A:Reference number: S23632; MUID:92146937
  A:Accession: S23632
  A:Status: preliminary
  A:Molecule type: mRNA
  A:Residues: 1-1010, 'V', 1012-1192, 'L', 1194-1274, 'I', 1276-1589 <DEC>
  A:Cross-references: EMBL:X63672; NID:g11056; PIDN:CA45211.1; PID:g11057
  C:Genetics:
  A:Gene: FlyBase:ph-p
  A:Cross-references: FlyBase:FBgn004861; FlyBase:FBgn004860
  A:introns: 12/2: 595/1: 745/2: 1340/1
  C:Superfamily: SAM homology
  C:Keywords: DNA binding; nucleus
  F:74-80,247-285,411-450,494-650,727-737,775-955,1032-1061/Region: glutamine-rich
  F:1510-1576/Domain: SAM homology <SAM>

```

```

alignment_scores:
  Quality: 132.00      Length: 537
  Ratio: 0.579        Gaps: 26
  Percent Similarity: 42.458      Percent Identity: 22.160

```

alignment_block:

US-09-303-518D-465 x T13606 ..

```

Align seg 1/1 to: T13606 from: 1 to: 1589
74 CAGATTGGCAACGATTCCTTTATCCGCGAGTTCTCGACCGTCAGCAT 123
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
955 GlnSerGlyGlnLeuGlnLeuSerValProPheSerValSerSerSe 971
124 TTGGAACCGGAGGGAATACCATTCGCGACGAGGAGGGAACCTTTC 173
   :::::::::::::::::::: |||||:::
971 rTThrThrProAlaGlyIleAlaThrSerSerAlaLeuGlnAlaLeuS 988
   :::::::::::::::::::: |||||:::
174 CGAGCCGACGGTCATATCGATTGGGAAACATACAAAGCCATCACTGG 223
   :::::::::::::::::::: |||||:::
988 erAlaSerGlyAlaIle.....PheGlnThrAlaLysPro..... 999

```

```

224 GCAACCTGTATCCAGCAGCGCGCCATTAAGAAATATCGGTCAT 273
   :::::::::::::::::::: ||| |||
1000 GlyThrCysSerSerSer.....SerProThrSe 1009
274 GTCCGCTTTCCGATCAGGCGCAGAAATCCATGCC.....CCTCGA 317
   ||||| ||||| ||||| ||||| ||||| |||||
1009 rSerAlaValThrIleThrAsnGlnSerSerThrProLeuValThrSer 1026
318 CAACCATG..... 325
1026 erThrValAlaSerIleGlnGlnAlaGlnThrGlnSerAlaGlnValHis 1042
325 ..... 325
1043 GlnHisGlnGlnLeuIleSerAlaThrIleAlaGlyGlyThrGlnGln 1059
326 .....CCTCAATTCGATTCGATTCGATTCGATTCGATTCGATTCG 362
1059 nProGlnGlyProProSerLeuThrProThrThrAsnProIleLeuAla 1076
363 TGACGATTCAGCCTTTACCGCATCCATTGGAGCGAGATACGAACCATC 412
   ::||| ||| :::::::::::::::::::: |||||:::
1076 erThr...SerMetMetAlaThrValGly...HisLeuSerThrAla 1090
413 CCGCCG.....ACGCTATGACGCGGCACAGGCG 441
   ||||| ||||| ||||| ||||| ||||| |||||
1091 ProProValThrValSerValThrSerThrAlaValThrSerProG 1107
442 GCGCGGTATCCCGCTCCCAAGGCGCA.....GGGATATATA 479
   ::||| ||| :::::::::::::::::::: |||||:::
1107 yGlnLeuValLeuLeuSerThrAlaSerSerGlyGlyGlySerIleP 1124
480 CAGCTACGACATTAAGGCGTTGCCCAAAATATCCGCTCAACCTGACCG 529
   ||||| ||||| ||||| ||||| ||||| |||||
1124 roIleAlaThr..... 1127
530 ACACCGCAGCAGCAGGACAAACGCGTTGCGACGCTTCCCAATACCGST 579
   ::||| ||| :::::::::::::::::::: |||||:::
1128 ThrLysGlnThrProSerLysGly...ProThrAlaThrLeuValPro 1143
580 AGTATGCTGACGACGAGTAGCGCAGGATTCGAAACGCGCACCCGATA 629
   ::||| ||| :::::::::::::::::::: |||||:::
1143 e.....GlySerProLysT 1148
630 CAGCCCGACCTGACAGATCGGSCAATGCCCGCAAGCTTCAACGGCA 679
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1148 hrProValSerGlyLysAspThrCysThrThrProLysSerSerThrPro 1164
680 CTGACGATTCGCA.....AAACATCATCGCGCGCGCAGGAGAAATT 723
   ::||| ||| :::::::::::::::::::: |||||:::
1165 AlaThrValSerAlaSerValGluAlaSerSer...SerThrGlyGluAla 1180
724 GTCGCGCAGCGCATGCGTCGACAGGATATACGCAAGGCTCAACATTCG 773
   ::||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1181 LeuSerAsnGlySerAlaSerAspArgSerSerThrProSerLys..... 1195
774 TGTATGACAGCGCTTGGTCTGCTTCCACCGCAAAACAAAGATGCGCGCA 823
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1196 .....GlyAlaThrThrProThrSerLysGlnSerAsn. 1206
824 TCAACGATTGGCAGATATGCGCACTCAAGACTATGCCGACGACGCC 873
1207 .....AlaAla 1208
874 ATCCGCGATTGGGACATGCCAAACCCCAATGCCGACAGCATAGAA... 921
   ::||| ||| :::::::::::::::::::: |||||:::
1209 ValGlnProProSerSerThrThrProAsnSerValSerGlyGlyGlu 1225
922 .....GCCGTACGACATATCT 937
1225 uProLysLeuAlaThrCysGlySerLeuThrSerAlaThrSerThrSer 1242
938 TTACGCGATCATCCCGTCAAAGGATTGAGAGCTGTTCGGGGAATAATC 987

```

```

1242 hrThrThrThrIle...ThrAsnGlyIleGlyValAlaAlaGlyThrAla 1257
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
988 GGCTTGGCGCGCATACGCGCATCTCTGTCAGCGGTGCGAGATGGCGCA 1037
      :||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1258 SerThrIleValSerThrIleAlaSerThrThrThrSer..... 1270
1038 GATCGCATTCGCGAAGGAAATCCGCGCTGCGACGAAATTTTCCGATG 1087
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1271 .....SerGlyThrPheThrThrS 1277
1088 CGCGATACGCCAAATACCCGCTCCCTTACCATTC...CGAATATCCGT 1134
      :||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1277 erCysThrSerThrThrThrThrThrThrSerSerIleSerAsnGlySer 1293
1135 TCAACCTTGAGCAGCGTTACGCGCAAGAAATATCACCCTCCCAACCGT 1184
      :||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1294 LysAspLeuProLysAlaMetIleLysProAsnValLeuThrHisValI 1310
1185 GCCGCGCTCAACGGAAGATGTGAACCTGCAACAAACGCGCACCGCA 1234
      :||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1310 e.....AspGlyPheIleIleGlnGlnAlaAsnGlnProPhePro 1324
1235 AGACCAAGCGCGCTTGACGCGTAAGGTTCCGAAATTTGAAAAAGAC 1284
      :||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1324 alThrArgLysArgTyr.....AlaAspLysAsp 1333
1285 GTAATATGAGATACGAGAAATTAATACCGCTGTACACAGTGAATCCCT 1334
      :||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1334 ValSer..... 1335
1335 AGATGAACCGCTCTTAATCTTAA..... 1359
      :||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1336 ..AspGlnPro.....ProLysLysLysAlaThrMetGlnLysAspI 1349
1360 .....GTTCTGTGCGATGCGCTCATCTGTGCTATACCTCCGAGA 1401
      :||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1349 IeLysLeuSerGlyIleAlaSerAlaProGlySerAspMetValAlaCys 1365
1402 ATTCAATACGCAAAATTACCAAGCAAGGTAGATC.....AGATA 1442
      :||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1366 GluGlnCysGlyLysMetGlnHisLysAlaLysLeuLysArgLysArgTyr 1382
1443 TATCCACCT 1452
      :||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1382 rCysSerPro 1385

```

seq_name: p1r2:S29605

seq_documentation_block:

glycoprotein 350/220 - human herpesvirus 4
 C:Species: human herpesvirus 4, Epstein-Barr virus
 C>Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 26-Aug-1999
 C:Accession: S29605
 R:Klein, K.; Mueller-Lantzsch, N.
 submitted to the EMBL data library, October 1992
 A:Description: Sequences of the membrane proteins gp 350/220 and p140 of Epstein-Barr vi
 A:Reference number: S29605
 A:Accession: S29605
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-886 <K1E>
 A:Cross-references: EMBL:X67776; NID:959163; PIDN:CAA47966.1; PID:959164
 C:Superfamily: Epstein-Barr virus membrane antigen gp350
 C:Keywords: glycoprotein

alignment_scores:

Quality: 131.50 Length: 588
 Ratio: 0.502 Gaps: 28
 Percent Similarity: 44.558 Percent Identity: 22.279

alignment_block:

US-09-303-518d-465 x S29605 ..
 Align seg 1/1 to: S29605 from: 1 to: 886

```

110 TCGACCGTACGATTTCCGACCCGACGGAAATACCACCTATTCG..... 154
      :||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
318 AsnThrThrAspIleThrIleValGlyAspAsnAlaThrIleThrSerValPr 334
155 .....GCAGCAGGGGGAACTTCCGAGCGCGACCGGCTCATATCGG 194
      :||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
334 oMetValThrSerGluAspAlaAsnSerProAsnValThrValThrAlaPr 351
195 ATTCG.....GAAACATACAAAGCCATCAGTTGGCGCA 226
      :||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
351 heThrAlaThrProAsnAsnThrGlnThrAspPheLysCysLysThrPhePr 367
227 ACCGTGTCATCCAGCAGCGCGCATTAAGGAATATTCGCTACATTCGTC 276
      :||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
368 LeuThrSerGlyThrProSerGlyCysGlnAsnIleSerGlyAlaPheAl 384
277 CGCT.....TTTCGATCAGCGGC..... 295
      :||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
384 aserAsnArgThrPheAspIleThrValSerGlyLeuGlyThrAlaProL 401
296 .....ACGAATCCATTCCTTCCGACACCATCCCTCAC 331
      :||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
401 yThrLeuIleIleThrArgThrAlaThrAsnAlaThrThrThrHis 417
332 ATTCGATTCGATGAGCCGCGTAGTCCCTTGACGCGATTCAGCTTAC 381
      :||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
418 LysValIlePheSerLys.....AlaProGlnSerThrThrHis 430
382 CGCATCCATTGGGACGATACGAACACCATCCCGCAGCGCTATGACG 431
      :||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
430 rSerProThrLeuAsnThrThrGlyPheAlaAlaProAsnThrThrThr 447
432 GCCACAGGCGCGGTATCCGCTCCCAAGGCGGAGGATATATACA 481
      :||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
447 Ly.....LeuProSerSerThrHisValProThr 456
482 GCTACGACATAAAGCGTTG.....CCGAAATATCCGCTCA..... 520
      :||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
457 AsnLeuThrAlaProAlaSerThrGlyProThrValSerThrHisAspVa 473
521 .ACGTGACGACAGCGAGACCGGACAGCGGTTTCACCGTTTCCA 569
      :||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
473 lThrSerProThrProAlaGlyThrThrSerGlyAlaSerProValThrPr 490
570 CAATACCGGTAGTATGCTGACGCAAGAGTAGGACGATTCAAACGCG 619
      :||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
490 roSerProSer.....ProArgAspAsnGlyThrGlnSerLysAla 503
620 CCACCCGATACGCCCCG...AGTGGACAGATCGGGCAATGCCCGCGAA 666
      :||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
504 ProAspMetThrSerProThrSerAlaValIleThrProThrProAsnAl 520
667 GCTTTCAACGCGACATGATTCGTCMAAAGATCATGCGCGCGCAGAG 716
      :||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
520 alThrSerProThrProAlaValIleThrProThrProAsnAlaThrSerP 537
717 AGAATTTGTCGCGCAGCGATCCGTGACAGGCTATTAAGCAAGCTCAA 766
      :||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
537 roThrLeuGlyLysThrSerProThrSerAlaValIleThrProThrPro 553
767 ACATTCGCTGTTATGCAAGCGCTTGCTGCTTCCACCGAAACAAGATG 816
      :||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
554 AsnAlaThrSerProThrProAlaValIleThrProThrProAsn..... 568
817 GCGCGCATCAAGATTTGGAGATATGCGCAACTCA..... 853
      :||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
569 .....AlaThrIleProThrLeuGlyLysThrSerProThrSerAlaVal 583
854 .....AAGACTATG 862

```

```

583 alThrThrProThrProAsnAlaThrSerProThrValGlyGluThrSer 599
863 CCGGACGACGATCCGGGATTT.....GGGACGTCGCAAAACCC 900
600 ProGlnAlaAsnThrThrAsnHisThrLeuGlyGlyThrSerSerThrPr 616
901 AATG.....CCGCACAAAG..... 913
616 ovalValThrSerProProLysAsnAlaThrSerAlaValThrThrGly 633
914 .....GCATAGAACCCGTCAGCAATATCTTACGGCAGCTCA 949
633 InHisAsnIleThrSerSerSerThrSerSerMetSerLeuArgProSer 649
950 TCCCGGTCGCAAGGATGAGCTGTGCGGGAATAACGCGTTGGCGGC 999
650 SerIleSer.....GluThrLeuSerProSerThrSerAspAsnSe 663
1000 ATCAGCGCAC.....ATCCTGTCAAGCGGTCGCGAGATGGCGAGATCGC 1043
663 ThrSerHisMetProLeuLeuThrSerAlaHisProThrIleGlyGlu 680
1044 ATTGCCGAAGGGAATCCGCGCTCA...GGCACAATTTGCGCATGGCG 1090
680 snIleThrGlnValThrProAlaSerThrSerThrHisValSerThr 696
1091 CATACGCCCAATATCCGCTCCCTTACATTCGCAATATCCGT..... 1134
697 SerSerProAlaProArgProGlyThrThrSerGlnAlaSerGlyProG 713
1135 ....TCMAACTGGAGCAGCGCTTACGCAAGAAACATCCGCTCTCA 1180
713 YAsnSerSerThrSerThrLysProGlyGluValAsnValThrLysGly 730
1181 CCGTCCGCGCTCAACGCAAGGAAGATGTGAACCTGGCAACCAACGCCAC 1230
730 hr...ProProLysAsnAlaThrSer..... 737
1231 CCGAAGACCAAAAGTCCGTTTACGCGTAAGGTTCCGAAATTTGAAA 1280
738 ProGlnAla.....ProSerGlyGlu 745
1281 AGACGTAAATATGATAGATGAGATTAATACCGCTGTACCAAGTAATC 1330
745 s.....ThrAlaValProThrValThrs 753
1331 CTAATAGATGACCCGCTTTAATCTTAAGGTTCTGTGGATCGGCTCAT 1380
753 eThrGlyGly.....LysAlaAsnSerThrThrGlyGlyLysHis 766
1381 TCCTGGTCATATACTGCAGAAATTCATACGCAAAATTTACCAAGCAAG 1430
767 ThrThrGlyHisGlyAlaArgThrSerThrGluProThrThrAspGly 783
1431 TAGAATCAGATATATCCACCTAAATTAATCTCTCTCAGCAGCGCTAC 1480
783 GlyAspSerThrThrProArgThrArgThrAsnAlaThrThrTyrLeu 800
1481 CAAAAGACCTAATATGATATTTGGATAAATTTGGTAATGAAGAGTACT 1530
800 roProSerThrSerSer.....LysLeuArgProArgTyrThr 812
1531 ....AAAGTCCATCAAGAACTAAAGTCAAGAATTTGAATGGAGTGTCA 1577
813 PheThrSerProProValThrThrAlaGlnAla.....ThrValPr 826
1578 ATTGCTAAACAGAGAGAGCA.....C 1603
826 ovalProThrSerGlnProArgPheSerAsnLeuSerMetLeuVal 843
1604 TTGATGGGCTACT 1617

```

843 euGlnThrProAlaSer 847

```

seq_name: p1r2:S50832
seq_documentation_block:
  atrophin-1 - human
  C:Species: Homo sapiens (man)
  C:Date: 14-Jul-1995 #sequence_revision 21-Jul-1995 #text_change 07-May-1999
  C:Accession: S50832
  R:Nagatani, S.; Yanagisawa, H.; Ohsaki, E.; Shiryama, T.; Tadokoro, K.; Inoue, T.;
  Nature Genet. 8, 177-181, 1994
  A:Title: Structure and expression of the gene responsible for the triplet repeat dlo
  A:Reference number: S50832; MUID:95144175
  A:Accession: S50832
  A:Status: preliminary
  A:Molecule type: mRNA
  A:Residues: 1-1184 <NAG>
  A:Cross-references: EMBL:D31840
  C:Genetics:
  A:Gene: GDB:DRPLA; B37
  A:Cross-references: GDB:270336; OMIM:125370
  A:Map position: 12p13.31-12p13.3112p-12p

```

```

alignment_scores:
  Quality: 131.50      Length: 684
  Ratio: 0.496        Gaps: 39
  Percent Similarity: 38.743  Percent Identity: 22.368

```

```

alignment_block:
  US-09-303-518d-465 x S50832 ..

```

```

Align seg 1/1 to: S50832 from: 1 to: 1184

```

```

65 CACAGCGCTGAGATTGGCAACGATCTTTATCCGAGCTTCTGAC 114
253 HisProPro.....ProThrThrProIleSerValSerSerGly 266
115 CGTCAGCATTTGCAACCCGACGGGAATACCACTTTCGCGACGAGGG 164
266 YAlaSerGlyAlaProProThrLysProProThrThrProValGlyGly 283
165 GGAACCTGCGGAGCGGCGGTCATATCGGATTTGGAAACATCAAGCC 214
283 LysAsnLeuProSerAlaProProProAlaAsnPhe.ProHisValThrPr 299
215 ATCAGTTGGCAACCTGTCATCCAGCAGG.CGGCAATTAAGAAATAT 263
299 oAsnLeuProProPro.....ProAlaLeuArgProLeuAsnAlaAs 314
264 CGGCTACA.....TTGTCGCTTTTCCGATCAGCGGCAAGAGTCC 304
314 eAlaSerProProGlyLeuGlyAlaGlnProLeuProGly..... 327
305 ATTCCCCCTTGACACACATGCGCTCACATTCGATTTGTGATGAAGCCGT 354
328 .....HisLeuPro..... 330
355 AGTCCCGTTGACGAGATTCAGCTTTTACCGCATTCATTTGGAGATTCGA 404
331 .....SerProTyr.....AlaMetGlyGlnGlyMetG 340
405 ACACATCCCGCGCGAGGCTATGACGGGCCACAGGGCGGCGATATCCG 454
340 IlyGlyLeuProProGlyProGlyLysGlyProThrLeuAlaProSerPro 356
455 .....CTCCCAAGGCGCGAGGATATATACAGCTACGACATAAAGGC 498
357 HisSerLeuProProAlaSerSerSerAlaProAlaProPro...MetAr 372
499 GTTGCCCAAAATATCCGCTCAACCTGACGACGACACGCGACACGGACA 548
372 gpheProTyrSerSerSerSerSerSerSerAlaAlaAlaSerSerSers 389

```



```

549 ACGGCTTGCACCGCTTTCACAAATACGGTAGTATGCTGACGCAAGAG 598
      ::::: |||::: |||::: |||
389 erSerSerSerSerSerSerLaserPrope..... 399
599 TAGGCGAGGATTCAAACGCCGCCACCGATACGCC...CCGACCTGGAC 645
      |||::: |||::: |||::: |||
400 .....ProLaserGlnAlaLeuProSerTyPr 409
646 AGATCGGCAATGCCGCCGAGCTTTCAACGGCACTGCGAGTATCGCA 695
      |||::: |||::: |||::: |||
409 ohSerSerProProPro.....ThSerLeuSerValSerSng 423
696 AAACATCATCGCGCGCGGCAAAATGTCGGCGACGCGATCCGCTGC 745
      ::::: |||::: |||::: |||
423 InProProLysTyThrGlnProSerLeuProSerGlnAlaVal..... 437
746 AGGTTATAGCGAAGGCTCAAAACATGCTGTATGACAGGCTTGGGCTG 795
      |||::: |||::: |||::: |||
438 .....TrpSerG 440
796 CTTTCCACCGAAACA.....AGATGGCGCGCATCAACGATT 833
      |||::: |||::: |||::: |||
440 nglProProProProProProTyGlyArgLeuAlaSer.... 455
834 GCGAGATATGGCGCACTCAAAAGCTATGCCGACGACGCGATCCGCGATT 883
      |||::: |||::: |||::: |||
456 .....AsnAlaHisProGlyProPheProProSerThrGly 467
884 GCGCATGCCAAACCCCAATGCCG.....CACAGGCATAGAA 921
      ::::: |||::: |||::: |||
468 AlaGlnSerThrAlaHisProProValSerThrHisHisHisHisG 484
922 GCGCTGCAATATCTTTACGCAATCATCCCGCAAGGATTTGAGC 971
      ::::: |||::: |||::: |||
484 InGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 500
972 TGTTCGGGAAATACGGCTTGGG..... 995
      |||::: |||::: |||::: |||
501 AsnSerGlyProProProProGlyAlaPheProHisProLeuGlnGly 517
996 ....CGGCATCACGCGCATCT..... 1014
      |||::: |||::: |||::: |||
517 ySerSerHisHisAlaHisProTyAlaMetSerProSerLeuGlySer 533
1014 ..... 1014
534 LeuArgProTyProProGlyProAlaHisLeuProProHisSerG 550
1015 .GTCAAGCGGTCCAGATGGCGGAGATGCCATTCGGAAGGAATCCG 1063
      |||::: |||::: |||::: |||
550 nValSerTySerGlnAlaGlyProAsnGlyProProValSerSerSers 567
1064 CC...GTCAAGCAATTTTCCGATGCGCATAC..... 1095
      ::::: |||::: |||::: |||
567 erAsnSerSerSerSerThrSerGlnGlySerTyProCysSerHisPro 583
1096 .....GCCAAATACCGCTCCCTTACCATT 1121
584 SerProSerGlnGlyProGlnGlyAlaProTyProPheProProVal 600
1122 CCGAATA.....TCCGTTCAACTGGACGAGGTACGGCA..... 1159
      |||::: |||::: |||::: |||
600 roThValThrThrSerSerAlaThrLeuSerThrValIleAlaThrVal 616
1160 .....AAGAAACATCACCTCTCAACCGTGGC 1188
      |||::: |||::: |||::: |||
617 AlaSerSerProAlaGlyTyTyTySerThrAlaSerProProGlyProPro 632
1189 CCGTCAACGGAAGAAATGTGAACCTGCAACAAACGCCACCG..... 1233
      |||::: |||::: |||::: |||
633 Pro.....TyGlyLysArgAlaProSerPr 641

```

```

1234 .....AAGACCAAGTCCGCTTTCACGTTAAGCGTTT..... 1266
      |||::: |||::: |||::: |||
641 oglyAlaTyTyTySerThrAlaThrPro.....ProGlyTyTyTyProG 655
1267 .....CCGATTTTGAAGAAAGAGTAATACATGACGAGA..... 1302
      |||::: |||::: |||::: |||
655 ySerProProSerPheArgThrGlyThrProProGlyTyTyTyGlyThr 671
1303 .....ATTAAATACCGCTGTACACAGTGA 1328
672 SerProProAlaGlyProGlyThrPheLysProGlySerProThrValG 688
1329 TCCTATAGATGAACCGCTTTATATCTTAAGT..... 1362
688 yProGlyProLeuProProAlaGlyProSerGlyLeuProSerLeuPro 705
1363 .....TCTGCGGATCGGCTCATTTGCTGTAACTGCGACA 1401
      ::::: |||::: |||::: |||
705 roProProAlaAlaProAlaSerGlyProProLeuSerAlaThr..... 719
1402 ATTCAATACGCAAAATTTACCAAGGCAAGGTACAGATATATATCCACC 1451
      ::::: |||::: |||::: |||
720 .....GlnIleLysGlnGlyProAl 726
1452 TAAATATAC...TCTCTTTCAGACCGCTTACCA.....AAGGACCTA 1492
      ::::: |||::: |||::: |||
726 agLysGlyTyGlyThrProGlySerProValProProAlaArgSerPro 743
1493 AT..... 1494
      ::::: |||::: |||::: |||
743 erProProProLysValValaSerValProSerHisAlaSerGlnSerAla 759
1495 .....AATGATATTGGATTA..... 1512
      |||::: |||::: |||::: |||
760 ArgPheAsnLysHisLeuAspArgGlyPheAsnSerCysAlaArgSerAs 776
1513 .....TTGGTAATGAATGACCTAA...GCTCCAT 1540
      |||::: |||::: |||::: |||
776 PLeuTyThrPheValProLeuGlnGlyProSerThrProArgSerGlyPro 793
1541 CA..... 1542
      ::::: |||::: |||::: |||
793 hTriTPArgArgCysGlyAlaGlyArgAlaAlaArgAlaArgArgLys 809
1543 .....AGACTAAAG 1553
810 GlyAlaArgAlaArgAlaGlyAlaGlyGlyGlyArgGlyGlyGly 826
1554 TCAGCAATTTGAATGGAGTTCATTTGCTTAAACAGAGAGA 1596
      ::::: |||::: |||::: |||
826 uArgAspValGlnArgSerValLysLeuAlaGlnGlyArg 840

```

seq_name: p1r2:T38459

seq_documentation_block:
 hypothetical divergent repeat-containing protein - fission yeast (Schizosaccharomyces
 C:Species: Schizosaccharomyces pombe
 C:Date: 20-Oct-2000 #sequence_revision 20-Oct-2000 #text_change 20-Oct-2000
 C:Accession: T38459, T38380
 R: Harris, D.; McDonald, S.; Barrell, B.G.; Rajandream, M.A.; Walsh, S.V.
 submitted to the EMBL Data Library, February 1996
 A:Reference number: 221794
 A:Accession: T38459
 A:Molecule type: DNA
 A:Residues: 1-1748 <HAR>
 A:Cross-references: EMBL:469368; PIDN:CA93290.1; GSPDB:GNO0066; SPDB:SPAC27F1.01c
 R:Klean, J.; Harris, D.; Barrell, B.G.; Rajandream, M.A.; Walsh, S.V.; Wood, V.
 submitted to the EMBL Data Library, September 1997
 A:Reference number: 221790
 A:Accession: T38380
 A:Molecule type: DNA
 A:Residues: 1457-1794 <MCL>
 A:Cross-references: EMBL:270691; PIDN:CA94638.1; GSPDB:GNO0066; SPDB:SPAC25G10.09c

A: Experimental source: clone c25610
C: Genetics:
A: Gene: SPAC27F1.01c; SPDB: SPAC25610.09c
A: Map position: 1

alignment_scores:
Quality: 131.50 Length: 520
Ratio: 0.632 Gaps: 25
Percent Similarity: 40.000 Percent Identity: 22.500

alignment_block:
US-09-303-518D-465 x T38459 ..

Align seg 1/1 to: T38459 from: 1 to: 1794

```
114 CCGTCAGCATTTTCGACCCGACGGGAAATACACCTATTTCGGACGAGG 163
|||||
1326 ProSerThrThrSerThrSerPheAsnThrAlaProIleProGlnGlnAl 1342
164 GGGAACTTCCGACGCGACGGTCATATCGGATTCGGAACATACATACG 213
1342 aProLeu.....GluAsnGlnPheSerIysM 1351
214 CATCAGTTGGGCAACCTGTCATCCAGCAGG.....CGGCA 250
|||||
1351 etSerLeuGlnProProValAlaArgProAlaValProThrSerProIysPro 1367
251 TTAAAGAAATATCGGCTACATTCGCTTCGATTCAGGCGACGAA 300
|||||
1368 GlnIleProAspSerSerAsnValHisAlaProPro.....Pr 1380
301 GTCATTCCTCCCTTCGACACCATGCT..CACATTCGATTCGTATGA 347
|||||
1380 oProValGlnProMetAsnAlaMetProSerHisAsnAlaValAsnAla 1397
348 ACCCGGTAGTCCGTTGACGATTCAGCCTTT..... 379
|||||
1397 rgProSerAlaProGlnArgArgAspSerPheGlySerValSerSerGly 1413
380 .....ACCGCATTCATTCGCGATTCGAAACCATCCG..... 415
|||||
1414 SerAsnValSerSerIleGlnAspGlnThrSerThrMetProLeuLysAl 1430
|||||
1430 aSerGlnProThrAsnProGlyAlaProSerAsnHisAlaProGlnVal 1447
450 TCCTCG.....CTCCCAAGGCGCGA 469
|||||
1447 alProProAlaProMetHisAlaValAlaProValGlnProLysAlaPro 1463
470 GGGATTTATACAGCTACGACATTAAGGCGTTGCCAAATATTCGCGCTC 519
|||||
1464 GlyMetValThrAsnAlaPro.....AlaProSe 1473
520 AACCTGACGACACCGACGACCGGACAGCGCTTGTGACCGTT.... 565
|||||
1473 rSerAlaProAlaProProAlaProValSerGlnLeuProProAlaValP 1490
566 .....TCACAATACCGGTAGTATCTGACGCAAGAG 598
|||||
1490 roAsnValProValProSerMetIleProSerValAlaGlnGln..... 1504
599 TAGCGACGATTCACAACGCGCCGACCGGATACGCGCCGAGCTGACAGA 648
|||||
1505 .....ProProSerSerValAlaProAlaThrAlaProSerSerThr... 1518
649 TCGGGCAATCCGCGAGCTTTCAACGCGCAGCTGAGATATGTCAAAAA 698
|||||
1519 .....LeuProProSerGlnSerSerPheAlaHisValProSerProA 1533
699 CATCATCGGGCGGCGAGGAAATTCGGCGGACGAGTACGCGTGCAGG 748
```

```
1533 la..... 1533
749 GTATATACGGAAGCTCAAAACATGTGCTTTATGACACGGCTGGTCTGCTT 798
1533 ..... 1533
799 TCCACCGAAMACAGATGCGCGCATCAAGATTTGGCAGATATGCGCGCA 848
|||||
1534 .ProPro..... 1535
849 ACTCAAAAGACTATGCCGACGACGATCCGGATTCGGGACATCCAAAC 898
|||||
1536 .....AlaProGlnHisProSerAlaAlaAlaLeuSerSerAla 1548
899 CCA.....ATCCGCGACAAGCA..... 916
|||||
1549 ProAlaAspAsnSerMetProHisArgSerSerProTyrAlaProGlnG 1565
917 .....TAGAACCGTCAGCAATATCTTTACGCGACATATCCCGCTCA 959
1565 uProValGlnLysProGlnAlaIleAsnAsnIleAlaProAlaThrAsnL 1582
960 AGGATTTGAGCGCTTCGGGAAATACGCTTGGG.....CGGCATCA 1003
|||||
1582 euLysThr.SerGlnSerPheSerProArgMetGlyProValAsnAsnSe 1598
1004 CGGCACATCTCTTCAGACGCGTCGACAGATGGCGAG.....ATCGCATTG 1047
|||||
1598 rGlySer.ProLeuAlaMetAsnAlaIaGlyGlnProSerLeuAlaVal 1614
1048 CCGAAGGGAATATCCCGCTGCGACGACATTTTCCGATTCGCGGCAATACC 1097
|||||
1615 ProAlaValProSerAlaProSerAsnHisPheAsnProPheAlaLysMe 1631
1098 CAATATCCGCTCCCT.....TACCATTCGCGAAATATCCGTTCA 1138
|||||
1631 tGlnProProAlaProSerProLeuGlnProSerGlnHisAspSerAspA 1648
1139 ACTTGAGCAGCGCTTACGCGAAGAAACATCACTCCCTCAACCGCGCG 1188
|||||
1648 snThrSerGlnHisGlyAspGlnGlnGlnLysAspSerGlnAspAspIle 1664
1189 CCGGACGCGAAGAAATGTGAACCTGGCAACCAACGCCACCGCAAGAC 1238
|||||
1665 ArgSerSerIysAspAlaAlaAlaLeuAlaLysLeu..... 1677
1239 CAAAGTCCGCTTTCGACGCTAAGGCTTCCGAATTTTGAAGAAAGACTCA 1288
|||||
1678 .....PheGlyGly..... 1680
1289 AATACGATACGAGATTAATATACGCTGTACACAGATGATCTATAGAT 1338
1681 .....MetAlaProAlaHisProValSer 1688
1339 GAACCCGCTTTAATCTTAAGGTTCTGTGATCGGCTCATTTCTTGTC 1388
|||||
1689 ThrProProValArgProGlnSerAlaAlaProProGlnMetSer..... 1703
1389 TATTAAGTCCAGAAATTAATACGCAAAATTAACCAAGCAAGATGATCA 1438
1704 .....AlaProThrProProProProPromets 1713
1439 GATATATCCGACCTAAATAATTAATCTTCCTTCAGCA...CCGCTACCAAA 1485
|||||
1713 erValProProPro.....ProSerAlaProPrometProAla 1725
1486 GGACCT 1491
|||||
1726 GlyPro 1727
seq_name: p1r2:G84693
```


A:Reference number: Z17898; MUID:98389687
 A:Accession: T14163
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 534-1122 <YUS>
 A:Cross-references: EMBL:AF027770; NID:93560502; PID:93560505; PIDN:AAC82548.1
 C:Genetics:
 A:Gene: exit
 C:Superfamily: Mycobacterium tuberculosis probable ABC transporter RV0194; ATP-binding C

alignment_scores:

Quality: 130.00 Length: 530
 Ratio: 0.568 Gaps: 31
 Percent Similarity: 38.814 Percent Identity: 23.380

alignment_block:

US-09-303-518D-465 x T14180 ..

Align seg 1/1 to: T14180 from: 1 to: 1122

```

69 CGCCTCAGATTGGCAACGATTCCTTTATCCGCGAGGTTCTCGACCGTC 118
   ||| :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
261 ATGAlaGluLeuGlySerArg.....GlyProGlyValGlyGlnProAr 275
   ||| :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
119 AGCATTTCGAACCCCGCGGGAATACCACTATT...CGCGACGAGGGGG 165
   ||| :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
275 gArSerArGProGluArGProArSProProLeuProArGLeuArGlyG 292
   ||| :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
166 GAACCTGGCGGCGGCGGTATCGGATTGGGAACATTCACAAGCCA 215
   ||| :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
292 lYProProArGArGArG..... 297
   ||| :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
216 TCAGTTGGCAACCTGTTTCATCCAGAGCGGCCATTAAAGGAATATCG 265
   ||| :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
298 .....GlyProProValGlyProGln.....ArgArSArThrAr 308
   ||| :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
266 G...CTACATTGTCGCTTTTCCGATCAGCGGACAGCAAGTCCATTCGCC 312
   ||| :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
308 gElYSerGluYsProValAlArGArGArGAlArG..... 320
   ||| :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
313 TTCGACAACCATGCTCACAATTCGATTCGTATGTAAGCGGTAGTCCGT 362
   ||| :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
320 ..... 320
   ||| :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
363 TGAAGGATTCAGCCTTTTACCGCATTCATTCGAGCGGATACGAACACATC 412
   ||| :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
321 ...ArglLeArGArGlyThrHisGlu.....ArgArGProAr 332
   ||| :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
413 CGCGCGAGCGCTATGACGGGCGCACAGGGCGG...CGGCTATCCGCTCC 459
   ||| :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
332 gCYsArGArGThrHisArGSerAlaLeuArGIlleSerAlaSerArGlyG 349
   ||| :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
460 AAAGGCGCGAGGATATATACAGCTACGACATAAAGGGATGGCCAAA 509
   ||| :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
349 lArGThrHisGlyAla.....GlyArGTrProAla 358
   ||| :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
510 TATCCGCTCAACCT...GACCGACACCGCGACACCGGACCAACGCTTG 556
   ||| :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
359 ProArGProArSProAlaGluArGAlaPro.....ArgThrProAlaAr 373
   ||| :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
557 TCGACCGTTTCACAAATACCGGTAGTGTGACGCAAGAGT..... 599
   ||| :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
373 gArGThrProProGlnLeuValAlGlnArGSpAlaGlyAsnGlnSerG 390
   ||| :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
600 .....AGGCGAGGATTCAAAGCGCGCACCCGATACAGCCCGAGCT 641
   ||| :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
390 lCYsHisArGAsnArGSerAlaAlaArGThrGlyLeuArG...ArgArG 405
   ||| :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
642 GGACAGATCGGCAATGCGCGCGAGCTTTCAGCGCACTGCGAGTATCG 691
   ||| :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
406 ArgThrThrGlyValAlArGArGArGArGArGArGArGProAlaProArG 422
   ||| :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

```

```

692 TCAAAAACATCATCGCGCGGACAGAGAAATTCGCGCGAGCGCA.... 737
   ||| :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
422 gSerArGProCYsArGArGArGlyAlaArGcysGluProValAlArGArGThr 439
   ||| :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
738 .....TGGCGTCAGGGGTATAGCGAAGGCTC 764
   ||| :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
439 hrProThrLeuArSProGlyProCYsAlaAlaGly..... 450
   ||| :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
765 AAACATTGCTGTATGACAGCGCTTGGGTCTGCTTCCACCGCAAAACAGA 814
   ||| :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
450 ..... 450
   ||| :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
815 TGCGCGCATACACGATTTGGCAGATGTGCGCACTCAAGACTATGCC 864
   ||| :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
451 .....ArgSerGlyArGProGlyAla.....Ala 459
   ||| :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
865 GCACGACCATCCCGCGATTTGGCGAGTCCCAAAACCCCAATCCGCGACAG 914
   ||| :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
459 rGSerArSPhis.....ArgArPArG 465
   ||| :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
915 CATAGAAGCCGT.....CAGCAATATCT 937
   ||| :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
466 HisArGHisArGAlaGlyHisArGArGArGHisThrProAlaAlaHisAl 482
   ||| :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
938 TTACGCGAGTCAT.....CCCGCTCAAGGAGATTGAGCTGTCCG 978
   ||| :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
482 aArGlyThrHisAsnGlyGlyProHisGln.....SerG 495
   ||| :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
979 GGAATAATACGGCTTGGCGGCA.....TCACGCGCATCTGTCAAGC 1021
   ||| :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
495 lYProAlaAlArGArGlyArGProGlyAlaAlaArGHisAlArGAlaSer 511
   ||| :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
1022 GGTGCGAGATGGGCGAGATGCGATTGGCGGAAGGAAATCCGCGTACG 1071
   ||| :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
512 GlyGlnArG..... 514
   ||| :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
1072 GACAATTTTCCGATGGCGCATACGCAATAC.....CGTCCCTTA 1115
   ||| :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
515 .....ArgHisThrProArGSerGlyArGArGProLeu 526
   ||| :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
1116 CCATTCCGGAATATCCGTTCAAACTTGAGCAGCGCTTACGCAAGAA 1165
   ||| :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
526 rGLeuProGlySerGlyAsp.....PrometlIleSerArSgluPro 539
   ||| :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
1166 ACATCACCTCTCAACCGTCCGCGGTCAACGGAAGAAATGTGAACATG 1215
   ||| :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
540 ValThrProProValLeuProValAlaThrAlaArGArSerGlyAlAr 556
   ||| :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
1216 G.....CAAAACACCGCACCGGACCAACCAAGTGGCGTTGACGG 1256
   ||| :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
556 pLeuIleAlaSpLeuArGlyArGArGThrAlaLeuAlaAlaValValT 573
   ||| :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
1257 TAAAGGTTTCCGAATTTGAAAAAGACGTAAATACGATACAGAAATTA 1306
   ||| :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
573 hrValGlyLeuAlaAla.....AlaAlaAlaSerVal 583
   ||| :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
1307 ATACCGCTGTACACAGATGATTCATAGATGACCCGCTTTAAATCT 1356
   ||| :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
584 ValProIleTyr.....Leu 589
   ||| :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
1357 AAAGTTCTGTGAGTCGCTCAT..... 1381
   ||| :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
589 uGlyMetLeuValAspArGValGlnHisGlyAspGlyThrArGlyLeuVal 606
   ||| :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
1382 .....CTTGGT 1387
   ||| :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
606 AlAlaLeuGlyThrValIleAlaLeuAlaAlaValAlaGlyGlyLeuGly 622
   ||| :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
1388 CTATTAATCGCAGAAATTCATACGCAAAATTTACCAAGCAAGTAAATC 1437
   ||| :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
623 ThrGlyLeuSerThrTyrLeuThrSerArGLeuGlyGlyGlnMetLeuAl 639
   ||| :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
1438 AGATATATTC.....CACTTAAATAATTATCTC 1463
   ||| :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

```

```

639 aaapleuArgValuArgValleuArgValleuAsnLeuProAlaThrL 656
1464 TCCTTGACGCGCTACCAAAAGACCTATATATGATTTTGATTAAT 1513
656 euleiGluSerGlyArgGlyAspLeuSerArgValGlyProAsp 672
1514 TTGGTATGATGACTAAAGTTCATCAAGACTAAAGTCAAGATTT 1563
673 ValAlaValAlaArgThrValAlaGlnValLeuProVal...IleLe 688
1564 GAATGGATTTCAATTGTCTAAACAGACAGACCACTTGATGGC 1613
688 uasnGlyPhePhe.....LeuGlyValValThrL 698
1614 TACTAGGATGTATGACATT 1633
698 euValGlyMetThrThrLeu 704

```

seq_name: p1r2:T51023

seq_documentation_block:

hypothetical protein B7F21.40 [imported] - Neurospora crassa
 C:Species: Neurospora crassa
 C:Date: 21-Jul-2000 #sequence_revision 21-Jul-2000 #text_change 21-Jul-2000
 C:Accession: T51023
 R:Schulte, U.; Alyn, V.; Hohelsel, J.; Brandt, P.; Fartmann, B.; Holland, R.; Nyakatura,
 submitted to the Protein Sequence Database, July 2000
 A:Reference number: 225286
 A:Accession: T51023
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-2649 <SCH>
 A:Cross-references: EMBL:AL389901; GSPDB:GN00116; NCSP:B7F21.40
 A:Experimental source: BAC clone B7F21; strain OR74A
 C:Genetics:
 A:Gene: NCSP:B7F21.40
 A:Map position: 6
 A:Insertions: 1619/3; 2584/1

alignment_scores:
 Quality: 130.00 Length: 593
 Ratio: 0.551 Gaps: 33
 Percent Similarity: 39.798 Percent Identity: 21.585

alignment_block:
 US-09-303-518d-465 x T51023 ..

Align seg 1/1 to: T51023 from: 1 to: 2649

```

236 TCACGACGCGCCCATTA.....AAGAAATATCGGCTAC 270
1961 SerThnGlnArgProMetProThrSerMetGlnGlnIlyrSerAlaAs 1977
271 ATTGTCCGCTTTTCGATCAGCGGACGACGATCCCTCCCTTGACAA 320
1977 nhrSerAla.....AlaGlnProValProProGlnIa 1989
321 CCATGCTTCATTCGATTCGTATGATGAGCCGTAGTCCCGTTGACGAT 370
1989 lAprProAlaIlePro.....ProGlnProLysLysValSer 2001
371 TCAGCCTTACC..... 382
2002 SerIlePheSerIleLeuAsnAspAsnProProAlaProAlaProAl 2018
383 .....GCATCCATT.....GGGACGATGAGA 404
2018 aProLysArgValAlaAsnAspValAlaSerMetProArgAlaIleSerTh 2035
405 ACACCATCCCGCCGACGGCTATGACGGCCAC..... 436

```

```

2035 eThrProProGlnGlnMetSerAlaArgProProGlnProProPro 2051
437 ...AGGGGGGGGCTATCCCGTCCCAAGGCGGAC.....G 471
2052 ProThrThrAlaValSerAlaSerGlnArgAspThrGlnAlaLeuG 2068
472 GATATATACAGCTACGACATTAAGGCGCTGCCCAAAATATCGGCT... 518
2068 lYrTySerTyAlaArgAsn.....ProProSer 2077
519 .....CAACTGACCGACACCGACGACCGGACGACGCGC 553
2078 AlaAlaGlnAlaAlaMetProSerLeuLysProThrIleThrGlnSer... 2093
554 TTGTGACCGCTTCCACATACCGGTATGCTGACGCAAGAGTACGC 603
2094 .....ProGlnProProIleMetAsnValAla..... 2102
604 GACGATTCAAACGCGCCGATACAGCCCGGAGGCTGACAGATCGGG 653
2103 ..ArgSerSerMetGlySerGlyMetGlnProGlnValSerAlaGly 2118
654 CAATGCGCGCCGAAAGCTTTCAAAGCGACTGCAGATATGCTCAAAACATCA 703
2119 AspAsnArgGluPheTySerArgHis...GlnTy...GlnProGlnH 2133
704 TCGGGCGCGGAGAGAAATGTCCGCGGACGATCCCGTCCAGAGTATA 753
2133 SGIProGlyAlaSerAsn..... 2139
754 AGCGAAGGCTCAACATTTGTTATGACGCGCTTGGGCTGCTTTCAC 803
2140 ..SerProValProHisGlnValHisIleTySerGlnSerAlaGlnHis 2155
804 CGAAACAGAGATGGCGCGCATCAACGATTTGGCAGATATGGCGCACTCA 853
2156 ProGlnGln.....HisLeuGln.GlnProGlnMetAlaTy...Pro 2168
854 AAGACTATGCGCGCAGACGACGCTCCGCGATGGGCGAG.....TCCAAAC 897
2169 SerGlnGlnGlnTyGlnProTyAlaThrSerGlnAlaLeuAlaAla 2185
898 CCCAATGCCGCGCACAAGCATAGAACCGCTCAGCAATCTTTCAGCGAGT 947
2185 rProThrProGlnIlyrAlaAlaHisProSerSerIleSerGlyArgArg 2202
948 CATCCCGCTCAAGAGGATGTGAGCTGTTCGGGCAAAATACGGCTGGCG 997
2202 lValAlaGlnSerAlaArgGlnIlyrPseSerThrGlnGlnGlnAla 2218
998 GCATCAGCGGACATCTGTCAAGCGGTCCGAGA..... 1030
2219 AlaValSerGlyLeuGlnGlnArgGlnGlnGlnGlnGlnGlnGln 2235
1030 ..... 1030
2235 nGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 2252
1031 .....TGGCGAGATCGCAT.....TGGCGAAG 1054
2252 lGlnGlnGlnGlnTyTrProProSerHisProThrProSerLysSerSerGln 2268
1055 GGAATCCGCGCGCAGGACAAATTTTCCAGTGCGGCGCATGCGCAATAC 1104
2269 AlaSerValProSerGlnIlyrAlaIleTrpAlaAlaGlnHisGlyProAsn.. 2284
1105 CCGTCCCTTACATTCGCCGAATATCCGTTCAACTTGAGACGAGCTTA 1154
2285 .....ValG 2286
1155 CGGCAAGAAACATCACCTCTCAACCGTGGCGCGCTCAACGGAAGA 1204
2286 lAlaIlyS.....ProProGlnMetGlySerAlaMetSerGlnGln 2299

```

```

1205 ATGTAAACTGGCAACAAACCCGACCCGAAAGCCAAAGTGGCGTTGAC 1254
      ::::::::::::::::::::
2300 GlnHisSerTyrGlnAlaThrProThrGlnGlnPro..... 2311
1255 GGTAAAGGTTCCGAAATTTTGAAGAAAGCCGTAATAATGATAGAGAT 1304
      ::::::::::::::::::::
2312 .....HISA 2313
1305 TAATACCGCTGTACCAAGTGAATCTATAGATGACCCGCTTTT... 1350
      ::::::::::::::::::::
2313 laLeuGlyLeu..ArgGluProAlaProArgGlyGlnAlaValPheSerA 2329
1351 .....AATCCTAAAGGTTCTGTGCGATCG...GCTATCT 1383
      ::::::::::::::::::::
2329 laHisGlyAlaGlnSerProThrGlySerValValSerHisGlnHis 2345
1384 TGGTGTATACCTGCGCAATTCATACGCAAAATATACCA..... 1422
      ::::::::::::::::::::
2346 ArgSerLeuAspGlyArgSerGlnPheProPromeProAspProArgAs 2362
1423 .....AGCAAGCTAGA..... 1434
      ::::::::::::::::::::
2362 parGlnAsnLeuArgArgGlyGluProValProProGlnGlnProT 2379
1435 ..ATCAGATATATC.....CCACCTAAATATCTCTCTCTCAGCA 1473
      ::::::::::::::::::::
2379 yValArgTyrValAsnThrProGlyProGlnHisGlyLysProGly 2395
1474 CCGCTACCAAGGACCTAATATGATATTTGGAT..... 1509
2396 ProProGlyGlnValProGlyArgGlyGlnAlaProAlaAspLeuArgme 2412
1510 .....AATTTGGTATATGATGACATAAGTCCATCAGACATA 1549
      ::::::::::::::::::::
2412 tGlnGlnMetSerGlnAlaArgSerTyrThrProGlyPro.....V 2426
1550 AAGCTCAAGAAATTTGAAATG..... 1569
      ::::::::::::::::::::
2426 aGlyValGlyPheGlnGlyMetGlyProProSerSerSerLeuGly 2442
1570 .....GATGTTCAATGTCATA 1586
2443 TyrProGlnGlnLeuArgAspAlaGlnLeuArgAspAlaGlnLeuArgG 2459
1587 AACAGAGAGAGACCACTTGA 1608
      ::::::::::::::::::::
2459 uMetGlySerArgAspLeuGly 2466
seq_name: pir2:T12970

```

```

seq_documentation_block:
  hypothetical protein T6H20.190 - Arabidopsis thaliana
  C:Species: Arabidopsis thaliana (mouse-ear cress)
  C:Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 22-Oct-1999
  C:Accession: T12970
  R:ChoiSue, N.; Robert, C.; Brottier, P.; Wincker, P.; Catolico, L.; Artiguenave, F.; Sa
  submitted to the Protein Sequence Database, July 1999
  A:Reference number: Z17586
  A:Accession: T12970
  A:Molecule type: DNA
  A:Residues: 1-510 <CHD>
  A:Cross-references: EMBL:AL096859; GSPDB:GN00061; ATSP:T6H20.190
  A:Experimental source: cultivar Columbia; BAC clone T6H20
  A:Gene: ATSP:T6H20.190
  A:Map position: 3
  A:introns: 139/3; 295/3; 318/3; 338/2

```

```

alignment_scores:
  Quality: 129.50      Length: 544
  Ratio: 0.514        Gaps: 26

```

```

Percent Similarity: 46.324      Percent Identity: 20.772
alignment_block:
  US-09-303-518D-465 x T12970 ..
  Align seg 1/1 to: T12970 from: 1 to: 510

```

```

242 AGCGCGCATTAAGAAATATGCGTACATTTGCGCTTTCCGATC 291
      ::::::::::::::::::::
23 ArgLysProArgLeuThrValTrrPalaLysGlnThrAlaPheGlnLeuG 39
292 GGGCAGCAAGTCCATCCCGCTTGACA..... 319
39 LysThrLysGlyAspAspSerGlnGlyLysGlnLysGlyLysAsn 56
      ::::::::::::::::::::
320 .....ACCATGCTTCACATTCGATTCGATGGAAG 349
      ::::::::::::::::::::
56 ropheGlnPheAspPheGlyLysLeuProAspMetLysSerLeuLeu... 71
350 CCGGTATC.....CCGTTGACGATTCAGCCTTACCGCATTCATGG 393
      ::::::::::::::::::::
72 ProValAlaThrAsnProSerThrGlyLeuValPheGlyAsnAsnArgly 88
394 GACGATACGACACATCCCGCGGACGCTATAGCGGCGCACGGCG 443
      ::::::::::::::::::::
88 sLysAspProGlyThrLePheValAlaGlyAlaThrngly..GlnAlaG 104
444 C.....GCTATCCCGCTCCAAAG 463
104 yLLeArgIleAlaGlnThrLeuLeuGlnArgGlyPheSerValArgAlaG 121
464 GCGGAGGATATA..TACAGTACGACATAAAGCGCTGCCAAAT 510
121 lValProAspLeuGlyAlaAlaGlnAspLeuAlaArgValAlaAlaThr 137
511 ATCCGCTCAACCTGACCGCAACCCGACGACCGCAACGCGCTTCGA 560
      ::::::::::::::::::::
138 TyrLysIleLeuSerAsnAspGlyValLysArgLeuAsnAlaValGlnse 154
561 CCGTTTCCACAAATACCGGTATGCTGACGCAAGGATGCGGACGAT 610
      ::::::::::::::::::::
154 rProPheGlnAspAlaGlnSer...IleAlaLysAlaIleGlyAsn... 168
611 TCAAAGCGGCGGACCGGATACAGCGCGGAGTGCAGATGGGCAATGCC 660
      ::::::::::::::::::::
169 .....AlaThrLysValValValAlaThrValGlyAlaThrGlnAsnGly 182
661 GCCGAGCTTTCAACGCGACCTGACAGATATGTCAAAACATCATCGCGCG 710
      ::::::::::::::::::::
183 ProAspAlaGlnValSerThrSerAspAlaLeu.....LeuValAlaG 197
711 GGCAGAGAAATTTGCGCGCAGCGCATGCCGTG.....CAGGTA 751
      ::::::::::::::::::::
197 nAlaAlaGlnLeuAlaGlyAlaSerHisValAlaIleValTyrAspLys 214
752 TAAAGGAGGCTCAACATTCGTTATGACGCGCTG..... 789
214 hrLleSerGlySerThrTyrAsnValLeuAspGlyIleThrSerPhe 230
790 GGTCTGCTTCCACCGAAACAGATGGCGCGCATCAACGATTTGGCAGA 839
      ::::::::::::::::::::
231 GlyAsnLeuPheAlaLysSerGlnProLeuThrLleSerAspLeuIleG 247
840 TATGGCGCACTCAAAACATATGCGCGACGACGATCCGCGATTTGGCAG 889
      ::::::::::::::::::::
247 ulysValAlaGlnThrAspValAlaTyrThrLeuIleLysThrSerLeu 264
890 TCCAA.....AACCCATGCGCGCAACAGCATAGAACCGTCAGCAAT 933
      ::::::::::::::::::::
264 hrGluAspPheSerProGlnLysAla..... 272
934 ATCTTACGCGCATCCCGCTCAAGGATGAGCTGTTGCGGGA 983
      ::::::::::::::::::::

```

273 ... TyrAsnValValValSerAlaGluGlySerAsnSer..... 284
 984 ATACGGCTTGGCGCATCAGCGACATCTCTCAAGCGGTGCGAGATGG 1033
 285 ... GlySerGlySerSerSerGluAlaTyrIys..... 295
 1034 GCGAGATCGCATTCGCCAAGGAAA..... TCCGCCGTGCGACAAAT 1077
 296 ValProLysLeuLysIleAlaSerLeuValAlaIleAspIle 308
 1078 TTTCGGCATGCGGCATACGCCAAATACCCCTCCCTACCATTTCCGAAA 1127
 309 PheAlaAsnThrAlaValAlaGlu..... 316
 1128 TATCGGTTCAAACTTGAGCAGCGTACGGCAAGAAAACATCAGCTCT 1177
 317 AsnLysValValGluValS 323
 1178 CAACCGTGCGG..... 1188
 323 erThrAspProSerAlaProSerArgProValAspGluLeuPheSerVal 339
 1189 ... CCGTCAACGGAAGAAATGTGAACCTGCGCAAC..... 1221
 340 IleProGluAspGlyArgArgLysValTyrAlaAspAlaIleAlaArgGly 356
 1222 AAACGCCACCGCAGACCAAGAGTCCGTTTGAC..... 1254
 356 uArgAlaGluGluGluAlaLysValAlaAlaAspLysAlaArgGluAla 373
 1255 GGTAAAGGTTTCGGAAT..... TTGAAAA 1281
 373 laGluAlaLysGluPheGluLysGluMetGluLysLeuSerGluLys 389
 1282 GACGTAATAATACGATACGAAATTAATCCGCTGACCAAGTGAATCC 1331
 390 GluAlaGluAlaAlaSerLeuAlaGluAspAlaGluGluLysAlaAspAl 406
 1332 TATAGAT..... GAACCGTCTTAATCCTAAAGTCTCTGCGGAT 1372
 406 aValGlyValThrValAspGlyLeuPheAsnLysAlaLysAspIleSerS 423
 1373 CGGCTCATTCCTGG..... 1386
 423 erGlyLeuSerTrpAsnLysLeuGlySerGlnPheAlaThrAlaIleGln 439
 1387 TCTAATACGCGCAATTAATCAATCAAGAAATTAACCAAGCAAG 1430
 440 AsnAlaSerGluThrProLysValGluValAlaThrValAlaArgGlyGlnAl 456
 1431 TAGAATCATATATATCCACCTAAATAATTACTCTCTCAGCAGCCGCTAC 1480
 456 aLysAlaIArgAsnLeuProPheLys..... LysAlaValValL 469
 1481 CAAAGACCACTAATATGATATTTGATTAATTTGTAATGAATGAGACT 1530
 469 ysglnArgProSerSerProPheAlaSerLysProLysGluGluArgPro 485
 1531 AAAGTCCATCAAGAACTAAAGTCAAGAAATTTGAATGGAGTGTCAATT 1580
 486 LysLysProGluLysGluValAlaArgLysValPheGlyGlyLeuPhe.... 500
 1581 GTCTAAACAGAGAGAGCAACTTGAT 1609
 501 LysGlnGluThrIleTyrIleAsp 508
 seq_name: pir2:s59310

seq_documentation block:
 probable membrane protein YMR317w - yeast (Saccharomyces cerevisiae)
 N/Alternate names: hypothetical protein YMR924.09
 C/Species: Saccharomyces cerevisiae
 C/Date: 29-Nov-1995 #sequence_revision 23-Feb-1996 #text_change 05-Nov-1999

C/Accession: S59310
 R/Churcher, C.M.
 Submitted to the EMBL Data Library, September 1995
 A/Reference number: S59302
 A/Accession: S59310
 A/Molecule type: DNA
 A/Residues: 1-1104 <CHU>
 A/Cross-references: EMBL:254141, GSPDB:GN00013, MIPS:YMR317w
 A/Experimental source: strain Ab972
 C/Genetics:
 A/Gene: MIPS:YMR317w
 A/Map position: 13R

alignment_scores:
 Quality: 129.50 Length: 430
 Ratio: 0.589 Gaps: 17
 Percent Similarity: 51.163 Percent Identity: 22.558

alignment_block:
 US-09-303-518D-465 x S59310 ..

Align seq 1/1 to: S59310 from: 1 to: 1104

86 ACAGATTCTTTATCCGGCAGGTTCTGACCGTCAAGATTTCGAACCCGAC 135
 180 ThrAlaValThrSerSerThrPheThrThr..... LeuThrAspValSer 194
 136 GGGAAATATCACCACTATTCGCGACAGGAGGGAATTCGCCAGCGCGAGCG 185
 194 rSerSerProLysIleSerSerSerGlySerAlaValThrSerValGlyT 211
 186 TCATATCGATTTGGGAAACATACAAAGCCATCAGTTGGGCAACCTGTGCA 235
 211 hrThrSerAsp..... AlaSerLysGluValPheSer 221
 236 TCACGAGCGGGGCCATTAAAGAA... ATATCGGCTACATGTCCGCTTT 282
 222 SerSerThrSerAspValSerSerLeuLeuSerSerThrSerProAl 238
 283 TCCGATACGCGGACGAGATCCATTCCTCCCTGACACACGCTCTCACA 332
 238 aserSerThrIleSerGluThrLeuProPheSerSerThrIle.....L 253
 333 TTCGATTTGTGATGAAGCCGATGCTCCGTGACGATTCAGGCTTTAC 382
 253 euserIleThrSerSerProValSerSerGluAlaProSerAlaThrSer 269
 383 GCATTCATTTGGACGATACGACACCATC..... 412
 270 SerSerValSerSerGluAlaSerSerSerThrSerSerValSerSe 286
 413 ...CCGCCAGCGCTATGACGGGCCACAGAGGGGGGCTATCCGCTC 458
 286 rglnAlaProLeuAlaThrSerSerValValSerSerGluAlaProSerS 303
 459 CAAAGCGCGAGGATATATACAGTACGACATTAAGGCGTGGCCCAAA 508
 303 erThrSerSerValValSerSerGluAlaProSerSerThrSerSer 319
 509 ATATCGCGCTCAACCTGACCGACACACCGACCGGACGAGGCTGTTC 558
 320 ValSerSerGluIleSerSerThrThrSerSerSerValSerSerGluAl 336
 559 GACCGTTTCACAAATACCGTAGTATGTCGACGCAAGAGTAGTGGGACGG 608
 336 aProLeuAlaThrSerSerValValSerSerGluAlaProSerSerThrS 353
 609 ATTCAAC.....GCGCCACCGCATACA..... 631
 353 erSerSerValSerSerGluAlaProSerSerThrSerSerSerValSer 369
 632GCCCGAGCTGACAGATCGGCGCATGCGCGGAGCTTTCAC 675

```

|||||  |||  :::::  :::::  |||  |||
370 SerGluAlaProSerSerThrSerSerValSerSerGluIleSerSe 386
676 GGCACCTGCAG.....ATATCGTCAAAAACATCATCGCGCGGAGAGA 719
:::  :::::  |||  |||  |||  |||  |||  |||
386 rThrLysSerSerValMetSerSerGluValSerSerAlaThrSerL 403
720 AATGTGCGCGGAGCGATGCGGTGAGGCTAAGCGAAGGCTCAACAA 769
:::  |||  :::::  |||  |||  :::  |||  :::::
403 euValSerSerGluAla...ProSerAlaIleSerSerLeuAlaSer 418
770 TGGCTGTATGACAGCGCTTGCGTCTTCCACGAAACAAAGATGGCG 819
|||  |||  |||  |||  |||  |||  |||  |||
419 ArgLeu.....PheSerSerLysAsnThrSera 428
820 CGCATCAACGATTGGCAGATATGGCGCACTCAAAAGACTATGCCGAC 869
:::  |||  |||  |||  |||  |||  |||  |||
428 ThrSerThrLeuValAlaThrGluAlaSerSerValThrSerSerLeu 445
870 AGCCATCCCGCAGTTGGCGAGTCCAAACCCCAATGCCGACAAAGCATAG 919
:::  |||  |||  |||  |||  |||  |||  |||
445 rGProSer.....SerGluThr..... 450
920 AAGCCGTCAGCAATATCTTTACGCGAGTCATCCCGTCAAAGGATTGGA 969
:::  |||  |||  |||  |||  |||  |||  |||
451 ...LeuAlaSerAsnSerIleIleGluSerSerLeuSerThrLys 466
970 GCTGTTCGGGGAAATATCGGCTTGCGGCGGATCA.....CGGC 1007
:::  |||  |||  |||  |||  |||  |||  |||
466 nSerThrValSerThrThrThrSerAlaIleSerSerThrLeuGlySerL 483
1008 ACATCTGTCAGACGCGTCGAGATGGCGAGATCGCATGCGCGAAG... 1054
:::  |||  |||  |||  |||  |||  |||  |||
483 yValSerSerSerAsnSerAlaGluMetAlaThrSerLysThrSerSerThr 499
1055 .....GGAATCCGCGCGCTGACG 1071
|||  |||  |||  |||  |||  |||  |||  |||
500 SerSerAspLeuSerLysSerSerValIlePheGlyAsnSerSerThra 516
1072 GACAAATTTCCGATCGCGCATACGCCAATATCCGCTCCCTTACCATTC 1121
|||  |||  |||  |||  |||  |||  |||  |||
516 ThrThrSerProSerAlaSerIleSerLeuThrAlaSer...ProLeu 532
1122 CCGAAATATCCGTTCAACTGTGAGCGAGCGCTTACGGCAAA..... 1161
||:  ||:  ||:  ||:  ||:  ||:  ||:  ||:
532 roSerVal.....TrpSerAspIleThrSerSerGluAlaSer 544
1162 .....GAAACATCACTCTCTCACCGCTGCGCGCTCAACGGGAAA 1202
:::  ||:  ||:  ||:  ||:  ||:  ||:  ||:
545 SerIleSerSerAsnLeuAlaSerSerSerAlaProSerAspAsnSer 561
1203 GAATGTGAACCTGGCAACAAACGCCACCGCAACCAAA 1242
:::  ||:  ||:  ||:  ||:  ||:  ||:  ||:
561 rThrIleAlaSerAlaSerLeuIleValThrLysThrLys 574

```

seq_name: pir2:E86185

seq_documentation_block:

hypothetical protein [imported] - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text-change 31-Mar-2001

C:Accession: E86185

R:Theologos, A.; Ecker, J.R.; Palm, C.J.; Federpiet, N.A.; Kaul, S.; White, O.; Alonso,

Chin, C.W.; Chung, B.K.; Com, L.; Conway, A.B.; Conway, A.R.; Cressy, T.H.; Dewar, K.;

ansen, N.F.; Hughes, B.; Huizari, L.

A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.

C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Lueros, J.S.; Matli, R.; Marziani,

Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.

A:Authors: Salberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,

ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.

A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.

A:Reference number: A86141; MUID:21016719

A:Accession: E86185
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-402 <STD>
A:Cross-References: GB:AE005172; NID:q2388584; PIDN:AA871465.1; GSPDB:GN00141
C:Genetics:
A:Map position: 1

alignment_scores: Quality: 129.00 Length: 313

Ratio: 0.791 Gaps: 9

Percent Similarity: 52.077 Percent Identity: 23.323

alignment_block:

US-09-303-518D-465 x E86185 ..

Align seq 1/1 to: E86185 from: 1 to: 402

```

98 TCCGCGAGTTCTGACCGTCACTTCGACCCGAGGAAATACACAC 147
|||  |||  |||  |||  |||  |||  |||  |||
72 SerSerSerThrLysSerAlaThrMetSerGlySerThrThrThr 88
148 CTATTCGCGACAGCGGGGAACTTCCGACGCGAGCGGTATTCGATT 197
|||  |||  |||  |||  |||  |||  |||  |||
88 rSerSerAlaThrAlaSerSerThrAlaSerThrThrSerSerLys 105
198 GGGAAACATACAAAGCCATCATGTTGGCGAAC..... 229
|||  |||  |||  |||  |||  |||  |||  |||
105 erThrSerThrSerThrSerThrLysThrLysThrThrMetThrLys 121
230 .....TGTTCATCCACAGCGCGCATTAAGAAATATCGGCTAC 270
|||  |||  |||  |||  |||  |||  |||  |||
122 ThrLysSerThrThrAlaSerAlaAlaProThrSerThrAlaSer 138
271 ATGTCCGCTTTTCCGATCAGCGGACGAGAGTCCATTCCTCCACAA 320
|||  |||  |||  |||  |||  |||  |||  |||
138 rThrSerSerLysThrSerThrSerThrSerThrLysThrLys 155
321 CCATGCGCTACATTCGATT...CTGATGAAGCGGATGCTGCGTACG 367
|||  |||  |||  |||  |||  |||  |||  |||
155 hrValThrGlySerThrIleGlyThrThrAlaSerAlaProThr 171
368 GATTACGCTTACCGCATCCATTCGAGGAGATACGACCATCCG... 415
|||  |||  |||  |||  |||  |||  |||  |||
172 ThrSerThrSerThrAlaAsnSerSerAlaSerSerThrThr 188
416 .....CCGACGGCTATGACGGGCGACAGCGGCGGCTATTC 452
|||  |||  |||  |||  |||  |||  |||  |||
188 rSerGlySerLysProThrAlaMetThrGlyThrThrAlaAsnThr 205
205 roSerAlaProThrSerSerProSerThrThrAsnSerSerThrAla 221
503 CCCAAATATCCGCTCAACCTGACG.....ACACCGC 537
|||  |||  |||  |||  |||  |||  |||  |||
222 AlaThrThrSerSerGlySerLysProThrThrValThrAlaThr 238
538 AGCAGCGCAACAGCGTTGCGACCGTTCCCAATAACCGGTATAGCT 587
|||  |||  |||  |||  |||  |||  |||  |||
238 asnThrSerSerSerAlaSerThrSerSerAlaSerProThr..... 252
588 GACCGAAGAGTAGCGACGATTCAAACGCCACCGCATACAGCCCGC 637
|||  |||  |||  |||  |||  |||  |||  |||
253 ..AsnSerSerThrSerThrProThrAsnSerSerAlaGlySerLys 268
638 .....ACGTGACAGATCGGCAATGCCCGGAAGCTTC 672
|||  |||  |||  |||  |||  |||  |||  |||
269 ThrThrMetThrGlyThrThrThrAsnThrSerSerThrThrThr 285
673 AACGCGCATCGAGATTCGTCAAACATCATCGCGCGGAGAGAAAT 722
|||  |||  |||  |||  |||  |||  |||  |||

```


285 rSeralaserThrThrLysSerSerSerSerSeralatThrAsnSers 302
723 TGTCGGCGAGGCGATCCGTGAGGGTAAAGCGAAGGCTCAACATG 772
302 erGly.....SerLysProSerThrLeu 309
773 CTGTATGACGGCTTGCTGCTTCCACCGAAMAAGATGGCGCG 822
310 SerThrThrThrAlaThrThrThrThrSerSerProThrAlaGlu 326
823 ATCAAGATTTGGCAGATATGGCCGACACACAAAGACTATGCCGAC 872
326 oSerThrThrThrAlaSerLysProAlaThrSerSerThrProProAla 343
873 CATCCGAGATTTGGGAGTCCCAAAACCCCAATGCCGACGAGATPAG 922
343 ro.ProThrThrThrAlaSerLysProArg...ArgThrHisAlaArg 358
923 CGGTCAACAATATCTTTACGGGAGTCATCCCGTCAA 959
358 rLeProGlnAsnThrPheHisThrArgIleArg 370
seq_name: p1r2:T49415

seq_documentation_block:
hypothetical protein BID4.300 [imported] - Neurospora crassa
C:Species: Neurospora crassa
C:Date: 02-Jun-2000 #sequence_revision 02-Jun-2000 #text_change 02-Jun-2000
C:Accession: T49415
R:Schulze, U.; Aign, V.; Hohnsiegel, J.; Brandt, P.; Fattmann, B.; Holland, R.; Nyakatura,
submitted to the Protein Sequence Database, May 2000
A:Reference number: 225022
A:Accession: T49415
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-634 <SCH>
A:Cross-references: EMBL:AL355928; GSPDB:GN00116; NCSP:BID4.300
A:Experimental source: BAC clone BID4; strain OR74A
C:Genetics:
A:Gene: NCSP:BID4.300
A:Map position: 6

alignment_scores:
Quality: 129.00 Length: 432
Ratio: 0.733 Gaps: 24
Percent Similarity: 40.741 Percent Identity: 21.759

alignment_block:
US-09-303-518d-465 x T49415

Align seg 1/1 to: T49415 from: 1 to: 634

84 AAACGATTTCTTTATCCGCGAGGTCGTGACCGTACGATTTGAAACCG 133
174 LysArgTyrPheGluProThrProThrAlaSerLeuProPheThrSer 190
134 ACGGGAATACCAACCATATCGGAGGAGGGAAGTCCGACGAGCGAC 183
190 rGlnThrGlnProLysProGlnAlaThrAlaGlySerAsnThrAspArg 207
184 GGTATATCGG.....ATTGGAAACATACAAAGCCATCAGTT 221
207 euglnTyrArgProIleThrLeuPheAspHisLeuLeuProThrLeu 223
222GGGCAACCTGTTTCATCCAGCAGG 244
224 HisArgGlnValSerValGlnLeuProSerGlnProValThrGluPro 240
245 CGGCATTAAGGAATATCGGCTACATGTGCCGTTTCCGATCAGCGG 294
240 rGlnGlnLeuArgThrLeuArg.....SerValArg 251

295 CACGAAGTCATTC...CCCTTCGACACACATCCTCATATCCGATTC 341
251 laArg.PheHisSerGluAspMetAspAlaHisGluValHisGluLys 267
342 TGATGAAGCCGGTAGTCCCGTTGACGAGATTCAGCCTTTACCGATTC 391
267 tAsnGlnLysAspAspProThrAsnValLeuAspLeu.SerThrProIle 283
392 GGGACGAGTACGAACACCATC..... 412
284 ProSerThrThrProSerPheLeuGlnGlyProGlySerGlnMetLeu 300
412 412
300 gTyrIleArgSerGluValLeuMetThrAlaValSerGlyPheSer 317
413 ..CGGCGGAGGCTATGACGGGCGACAGGGCGGGCTATCCGCTCCCA 460
317 erProProAlaSerProThrValValLysSerGlyLeuValThrPro 333
461 AAG.....GGGAGGAGATA..... 475
334 GlyAspThrArgLysGlyThrLeuPheProGlnGlyValSerThrPro 350
476TATACAGCTACGACATAAAGCGGTCGCCAATAATATCCGCT 518
350 uLeuGluThrThrAlaThr.....ProGluProLysSer 362
519 CAACCTGACCG..... 529
352 snThrLeuProPheProAsnLeuThrValAsnAsnArgAsnPheThr 378
530ACAACGCGACGACGCGACGAGCGCTTGTCGACGCTTC 567
379 AspProThrPheIleSerThrSerProProAsnLeuIleProSerPro 395
568 CACAATACCGGCTAGTATGCTGACGCAAGAGTAGCGAGATTCAAACG 617
395 aProAlaPro.....Asn 400
618 CGCCACCGGATACAGCCCGGAGCTGGACAGATCGGCAATGCCGCGAG 667
400 hrProProAsnThrThrProSer...LysLysArgGlyArgProPro 415
668 CTTCACAGCGCAGTCGAGATATGCTCAAAAACATCATCGGCGGCGAGA 717
416 SerLysSer.....HisSerLysSerProSerSerLysGly 428
718 GAATTTGTGCGCGCAGCGAGATGCCGTGACGGTATAAGCAGAGCTCAA 767
428 nArgIleThrAlaThrThr..... 434
768 CATGCTGTATGACAGGCTTGCGTCTTCCACCGAAA.....ACA 811
435ProGlnLysArgGlnGlu 440
812 AGATGGCGCGCATCA...ACGATTTGGCAGATAGGGCGCAAC.....TC 852
441 ArgLeuAlaGalaSerProSerSerThrProGlnThrThrAsnValGln 457
853 AAAGACTATGCCGAGCAGCAGCCATCCGAGATTTGGCAGCTCAAAACCC 902
457 rTrpSerLeuLeuSerGln.....GlyGlnHisHisThrPro. 469
903 TGCCGACACAGGCGATAGAGCCGTCACA..... 931
470 ..ProHis.....CysAspAlaPheGlyPheAspThrTrpIle 481
932ATATCTTACGCGACATC 948
482 ThrValGlnGlnGlyMetValGlyPheGlyTrpPheGlyThrAlaSer 498
949 ATCCCGGTCAAGAGGATTTGAGGTGTCGGGGAATAATACGCTTGCGCG 998


```

321 .....CAAGGATTCCTATAGATGAACCCGCTTTAATCTTAAAGGTTG 1364
      :: ::||| :: ::::::::::::::: ||
237 ThrGlnSerSerProLaserThrLeuAlaTyrSerProSerValSe 1253

```

